(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 August 2004 (05.08.2004)

PCT

(10) International Publication Number
WO 2004/065401 A1

(51) International Patent Classification[7]: C07H 15/203, 15/26, 15/10, C07C 229/30, 237/16, C12P 19/44, A61K 31/7032, A61P 31/00

(21) International Application Number:
PCT/CA2004/000068

(22) International Filing Date: 21 January 2004 (21.01.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/441,123	21 January 2003 (21.01.2003)	US
60/469,810	13 May 2003 (13.05.2003)	US
60/491,516	1 August 2003 (01.08.2003)	US
60/494,568	13 August 2003 (13.08.2003)	US

(71) Applicant *(for all designated States except US)*: **ECOPIA BIOSCIENCES INC.** [CA/CA]; 7290 Frederick-Banting, Saint-Laurent, Québec H4S 2A1 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BACHMANN, Brian, O.** [US/US]; 2905 Snowden Rd., Nashville, TN 37204 (US). **MCALPINE, James, B.** [US/CA]; 40 Upper Trafalgar Place, Montreal, Québec H3H 1T3 (CA). **ZAZOPOULOS, Emmanuel** [CA/CA]; 455 Saint-Pierre, apt. 410, Montreal, Québec H2Y 2M8 (CA). **FARNET, Chris, M.** [US/CA]; 772 Bloomfield, Outremont, Québec H2V 3S3 (CA).

(74) Agent: **LOOPER, Ywe, J.**; 7290 Frederick-Banting, Saint-Laurent, Québec H4S 2A1 (CA).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i)) for all designations*

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *with international search report*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

— *with (an) indication(s) in relation to deposited biological material furnished under Rule 13bis separately from the description*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE AS A PHARMACEUTICAL

(57) Abstract: This invention relates to a new class of polyene polyketides, their pharmaceutically acceptable salts and derivatives, and to methods for obtaining the compounds. One method of obtaining these compounds is by cultivation of novel strains of *Streptomyces aizunensis*; another method involves expression of biosynthetic pathway genes in transformed host cells. The present invention further relates to the novel strains of *Streptomyces aizunensis* used to produce these compounds, to the use of these compounds and their pharmaceutically acceptable salts and derivatives as pharmaceuticals, in particular to their use as inhibitors of fungal cell growth and cancer cell growth. The invention also relates to pharmaceutical compositions comprising these novel polyketides or a pharmaceutically acceptable salts or derivatives thereof. Finally, the invention relates to novel polynucleotide sequences and their encoded proteins, which are involved in the biosynthesis of these novel polyketides.

WO 2004/065401 A1

TITLE OF INVENTION: POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE AS A PHARMACEUTICAL

RELATED APPLICATIONS:

This application claims priority to U.S. Provisional Application 60/441,123 filed January 21, 2003; U.S. Provisional Application 60/494,568 filed August 13, 2003; U.S. Provisional Application 60/469,810 filed May 13, 2003; and U.S. Provisional 60/491,516 filed August 1, 2003.

FIELD OF INVENTION:

This invention relates to a new class of polyene polyketides, their pharmaceutically acceptable salts and derivatives, and to methods for their production. One method of obtaining these novel polyketides is by cultivation of novel strains of *Streptomyces aizunensis;* another method involves expression of the biosynthetic gene cluster of the invention in transformed host cells. The compounds may also be produced by known strains of certain bacteria. The invention also encompasses the novel strains of *Streptomyces aizunensis* which produce these compounds, as well as the gene cluster which directs the biosynthesis of these compounds. The invention also includes the use of these novel polyketides and their pharmaceutically acceptable salts and derivatives as pharmaceuticals, in particular, to their use as inhibitors of fungal and bacterial cell growth, inhibitors of cancer cell growth and for lowering serum cholesterol and other steroids. The invention also encompasses pharmaceutical compositions comprising these novel polyketides, or pharmaceutically acceptable salts or derivatives thereof.

BACKGROUND:

Actinomycetes comprise a family of bacteria that are abundant in soil and have generated significant commercial and scientific interest as a result of the large number of therapeutically useful antibiotics, antifungals, anticancer and cholesterol-lowering agents, produced as secondary metabolites by these bacteria. Many actinomycetes, particularly those of the Streptomyces genus,

have been extensively studied because of their ability to produce a notable diversity of biologically active metabolites. The intensive search for new natural products has led to the identification of new species of bacteria and the creation of improved strains.

Polyene polyketides are a group of natural products produced by actinomycetes that have generated significant commercial interest. For example Sakuda *et al*, 1996 *J. of Chem. Soc., Perkin trans.* 1, 2315-19; and Sakuda *et al.*, *Tetrahedron Letters*, Vol 35, No. 16, 2777-2789 (1995) disclose the linear polyene linearmycin A produced by a Streptomyces sp. Sakuda *et al.* report that linearmycin A has shown both antifungal and antibacterial activity. Pawlak *et al. J of Antibiotics*, Vol. XXXIII No. 9, 989-997 disclose the polyene macrolide lienomycin produced by *Actinomyces diastatochromogenes*. Pawlak *et al.* report that lienomycin has shown antifungal, antibacterial and anti-tumor activity. Antifungal activity of polyene macrolides has also been correlated with hyperchlesterolemic effect (C.P. Schaffner, Polyene Microlides in Clinical Practice, in Macrolide Antibiotics: Chemistry, biology and practice, S. Omura, ed. Academic Press (1984), p. 491; C.P. Schaffner and H.W. Gordon, *Proc Natl. Acad. Sci. U.S.A.* **61**, 36 (1968)).

Polyketides have carbon chain backbones formed of two-carbon units through a series of condensations reactions and subsequent modifications. Type I polyketides are synthesized in nature by modular polyketide synthase (PKS) enzymes having a set of separate catalytic active sites for each cycle of carbon chain elongation and modification. Because of the multimodular nature of PKS proteins, much is known of the specificity and mechanism of the biosynthesis of polyketides.

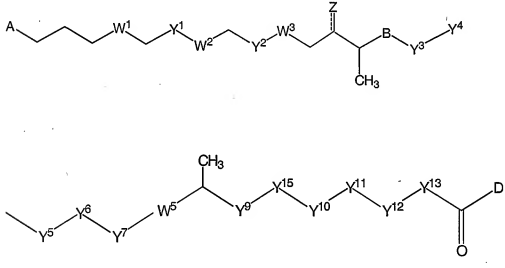
Although many biologically active compounds have been identified, there remains the need to obtain novel naturally occurring compounds with enhanced properties. Current methods of obtaining such compounds include screening of natural isolates and chemical modification of existing compounds, both of which are costly and time consuming. Current screening methods are based on general biological properties of the compound, which require prior knowledge of the structure of the molecules. Methods for

chemically modifying known active compounds exist, but still suffer from practical limitations as to the type of compounds obtainable.

Thus, there exists a considerable need to obtain pharmaceutically active compounds in a cost-effective manner and with high yield. The present invention solves these problems by providing improved strains of *Streptomyces aizunensis* capable of producing potent new therapeutic compounds, as well as reagents (*e.g.* polynucleotides, vectors comprising the polynucleotides and host cells comprising the vectors) and methods to generate novel compounds by de novo biosynthesis rather than by chemical synthesis.

SUMMARY OF THE INVENTION:

The present invention encompasses compounds of Formula I:

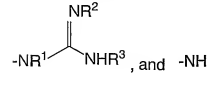


Formula I

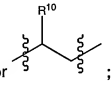
and pharmaceutically acceptable salts thereof;

wherein,

A is selected from the group consisting of $-NR^1R^2$, $-N=CR^1R^2$,



R^1, R^2, R^3 and R^4 are each independently selected from the group consisting of H, C_{1-6} alkyl, C_{2-6} alkenyl, C_{3-6} cycloalkyl, C_{2-6} heterocycloalkyl, aryl, heteroaryl and amino acid, wherein said alkyl, alkenyl, aryl and heteroaryl are optionally substituted with a group selected from halogen, OH, NO_2, NH_2 or aryl, said aryl being optionally further substituted with one or more groups independently selected from halogen, OH, NO_2 or NH_2;

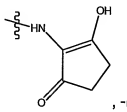
B is selected from ethene-1,2-diyl or 

;

wherein R^{10} is oxo or OR^{11};

wherein R^{11} is H or a heterocycloalkyl, the heterocycloalkyl being optionally substituted with 1-4 substituents selected from OX, C_{1-3} alkyl and $-O-C(O)R^1$, wherein X is H or, when there are at least two neighboring substituent groups that are OX, then the X can be a bond such that the two neighboring oxygen groups form a five-membered acetal ring of the formula:

; wherein R^5 and R^6 are each independently selected from the group consisting of H, C_{1-6} alkyl, and C_{2-7} alkenyl;

D is selected from 

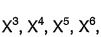
, $-NR^{12a}R^{12a}$, and OR^{12}, wherein

R^{12} is selected from H and C_{1-6} alkyl optionally substituted with 1 to 2 phenyl groups, wherein the phenyl group is optionally substituted with C_{1-6} alkyl or halo;

R^{12a} and R^{12a} are each indepedently selected from H, C_{1-6} alkyl, C_{2-6} alkenyl, C_{3-6} cycloalkyl, C_{2-6} heterocycloalkyl, aryl, heteroaryl and amino acid, wherein said alkyl, alkenyl, aryl and heteroaryl are optionally substituted with a group selected from halogen, OH, NO_2, NH_2 or aryl, said aryl being optionally further substituted with one or more groups independently selected from halogen, OH, NO_2 or NH_2;

W^1 is  ;

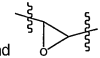
W^2 is ;

W^3 is  ;

W^5 is  ;

X^1, X^2, X^3, X^4, X^5, X^6, X^7, X^8, X^9, X^{12} and X^{13} are each independently selected from H, -C(O)-R^7 and a bond such that when any of two neighboring X^1, X^2, X^3, X^4, X^5, X^6, X^7, X^8, X^9, X^{12} and X^{13} is a bond then the two neighboring oxygen atoms and their attached carbon atoms together form a six-membered acetal ring of the formula:

R^5, R^6 and R^7 are each independently selected from H, C_{1-6} alkyl, C_{2-7} alkenyl;

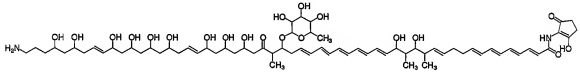
Y^1, Y^2, Y^3, Y^4, Y^5, Y^6, Y^7, Y^9, Y^{10}, Y^{11}, Y^{12}, Y^{13} and Y^{15} are each independently selected from the group consisting of ethene-1,2-diyl, ethane-1,2-diyl and  ; wherein said ethene-1,2-diyl and ethane-1,2-diyl groups are optionally substituted with a methyl group;

Z is selected from OH, NHR^8,  and when the dotted line is a bond then Z is oxo, or NR^9;

R^8 is selected from H, C_{1-6} alkyl, C_{2-6} alkenyl;

R^9 is C_{1-6} alkyl optionally substituted with aryl.

The invention is also directed to the Compound 2(a), a linear glycosylated polyketide with an amidohydroxycyclopentenone component, and pharmaceutically acceptable salts thereof:



Compound 2(a)

The systematic name for Compound 2(a) has been determined to be: 56-Amino-15,17,33,35,37,41,43,45,47,51,53-undecahydroxy-14,16,30-trimethyl-31-oxo-29-(3,4,5-trihydroxy-6-methyl-tetrahydro-pyran-2-yloxy)-hexapentaconta-2,4,6,8,12,18,20,22,24,26,38,48-dodecaenoic acid (2-hydroxy-5-oxo-cyclopent-1-enyl)-amide.

The invention encompasses pharmaceutical compositions of compounds of Formula I comprising, a therapeutically effective amount of the

compound of Formula I or a pharmaceutically acceptable salt thereof, and a pharmaceutically acceptable carrier. In particular, the invention is directed to pharmaceutical compositions of compound 2(a) comprising, a therapeutically effective amount of the compound 2(a) or a pharmaceutically acceptable salt thereof, and a pharmaceutically acceptable carrier.

The present invention is also directed to methods for producing the compound 2(a) and related compounds, including compounds of Formula I and Formula II as defined herein. Such methods comprise the steps of cultivating cells derived from a *Streptomyces aizunensis* strain, incubating said cultured cells aerobically in a growth medium for such time as is required for production of the desired compound, extracting said medium with a solvent such as methanol or ethanol and purifying the compound from the crude extract. The *Streptomyces aizunensis* strain which may be used in the methods of the invention may be NRRL B-11277 or a mutant thereof. A preferred strain of *Streptomyces aizunensis* useful in the methods of the invention is a mutant strain identified as [C03]023 (deposit accession number IDAC 070803-1); a most preferred strain of *Streptomyces aizunensis* useful in the methods of the invention is a mutant strain identified as [C03U03]023 (deposit accession number IDAC 231203-02). The invention also encompasses the *Streptomyces aizunensis* strains identified by deposit accession numbers IDAC 070803-1 and IDAC 231203-02.

The invention also includes methods of inhibiting fungal cell growth, which comprise contacting a fungal cell with a compound of Formula I, a compound of Formula II or compound 2(a), or a pharmaceutically acceptable salt thereof. In addition, the invention encompasses methods for treating a fungal infection in a mammal, which comprise administering to a mammal suffering from such an infection, a therapeutically effective amount of a compound of Formula I, a compound of Formula II or compound 2(a), or a pharmaceutically acceptable salt thereof. The methods of the invention are particularly useful for treating fungal infections or inhibiting the growth of fungal cells in mammals caused by *Candida albicans.* The invention also encompasses methods for treating or inhibiting other types of fungal infections in a subject, wherein said fungal infections include those caused by *Candida*

sp. such as *C. glabrata, C. lusitaniae C. parapsilosis, C. krusei, C. tropicalis, S. cerevisiae; Aspergillus sp.* such as *A. fumigatus, A. niger, A. terreus, A. flavus; Fusarium spp.; Scedosporium spp.; Cryptococcus spp.; Mucor ssp.; Histoplasma spp.; Trichosporon spp.; and Blaspomyces spp.* Such methods comprise administering to a subject suffering from the fungal infection, a therapeutically effective amount of a compound of Formula I, Formula II or compound 2(a), or a pharmaceutically acceptable salt thereof.

The invention also provides methods of inhibiting cancer cell growth, which comprise contacting said cancer cell with a compound of Formula I, Formula II or compound 2(a), or a pharmaceutically acceptable salt thereof. The invention further encompasses methods for treating cancer in a subject, comprising administering to said subject suffering from said cancer, a therapeutically effective amount of a compound of Formula I, Formula II or compound 2(a) or a pharmaceutically acceptable salt thereof. Examples of cancers that may be treated or inhibited according to the methods of the invention include leukemia, non-small cell lung cancer, colon cancer, CNS cancer, melanoma, ovarian cancer, renal cancer, prostate cancer and breast cancer.

The present invention also provides the biosynthetic locus from *Streptomyces aizunensis* (NRRL B-11277) which biosynthetic locus is responsible for producing the compound of Formula 2(a). *Streptomyces aizunensis* was not previously reported to produce Compound 2(a). We have now discovered, in the *Streptomyces aizunensis* genome, the gene cluster responsible for the production of the Compound 2(a). Thus the invention provides polynucleotides and polypeptides useful in the production and engineering of compounds of Formula I and Compound 2(a). The invention also provides chemical modifications of compounds of Formula I and Compound 2(a).

In one aspect, the invention relates to the biosynthetic locus for production of a polyketide of Formula I and provides, in one embodiment, an isolated, purified or enriched nucleic acid for production of a polyketide of Formula I comprising a nucleic acid encoding at least one domain of the

polyketide synthase system formed by the polyketide synthases of SEQ ID NOS: 21, 23, 25, 27, 29, 31, 33, 35 and 37.

In a further embodiment, the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 21 and comprises a nucleic acid selected from the group consisting of: a) SEQ ID NO: 22; b) the nucleic acid of residues 169-354 of SEQ ID NO: 22, the nucleic acid of residues 421-1698 of SEQ ID NO: 22, the nucleic acid of residues 1789-3093 of SEQ ID NO: 22, the nucleic acid of residues 3910-4551 of SEQ ID NO: 22, the nucleic acid of residues 4807-4992 of SEQ ID NO: 22, the nucleic acid of residues 5068-6354 of SEQ ID NO: 22, the nucleic acid of residues 6403-7686 of SEQ ID NO: 22, the nucleic acid of residues 8497-9135 of SEQ ID NO: 22, the nucleic acid of residues 9388-9573 of SEQ ID NO: 22, the nucleic acid of residues 9643-10920 of SEQ ID NO: 22, the nucleic acid of residues 10978-12267 of SEQ ID NO: 22, the nucleic acid of residues 12304-12624 of SEQ ID NO: 22, the nucleic acid of residues 13834-14487 of SEQ ID NO: 22, the nucleic acid of residues 14731-14916 of SEQ ID NO: 22, the nucleic acid of residues 15019-16314 of SEQ ID NO: 22, the nucleic acid of residues 16378-17649 of SEQ ID NO: 22, the nucleic acid of residues 18439-19080 of SEQ ID NO: 22, the nucleic acid of residues 19330-19515 of SEQ ID NO: 22, the nucleic acid of residues 19585-20862 of SEQ ID NO: 22, the nucleic acid of residues 20935-22206 of SEQ ID NO: 22, the nucleic acid of residues 23107-23754 of SEQ ID NO: 22, the nucleic acid of residues 24004-24189 of SEQ ID NO: 22; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 23 and comprises a nucleic acid selected from the group consisting of: a) SEQ ID NO: 24; b) the nucleic acid of residues 109-1386 of SEQ ID NO: 24, the nucleic acid of residues 1477-2757 of SEQ ID NO: 24, the nucleic acid of residues 2794-3114 of SEQ ID NO: 24, the nucleic acid of residues 4231-4881 of SEQ ID NO: 24, the nucleic acid of residues 5116-5301 of SEQ ID NO: 24, the nucleic acid of residues 5380-6645 of SEQ ID NO: 24, the nucleic acid of residues 6694-7977 of SEQ ID

NO: 24, the nucleic acid of residues 8878-9519 of SEQ ID NO: 24, the nucleic acid of residues 9772-9957 of SEQ ID NO: 24; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 25 and comprises a nucleic acid selected from the group consisting of: a) SEQ ID NO: 26; b) the nucleic acid of residues 106-1383 of SEQ ID NO: 26, the nucleic acid of residues 1447-2721 of SEQ ID NO: 26, the nucleic acid of residues 2755-3081 of SEQ ID NO: 26, the nucleic acid of residues 4315-4965 of SEQ ID NO: 26, the nucleic acid of residues 5206-5391 of SEQ ID NO: 26, the nucleic acid of residues 5491-6768 of SEQ ID NO: 26, the nucleic acid of residues 6841-8142 of SEQ ID NO: 26, the nucleic acid of residues 8941-9582 of SEQ ID NO: 26, the nucleic acid of residues 9832-10017 of SEQ ID NO: 26, the nucleic acid of residues 10081-11358 of SEQ ID NO: 26, the nucleic acid of residues 11407-12675 of SEQ ID NO: 26, the nucleic acid of residues 13480-14118 of SEQ ID NO: 26, the nucleic acid of residues 14383-14568 of SEQ ID NO: 26, the nucleic acid of residues 14638-15912 of SEQ ID NO: 26, the nucleic acid of residues 15967-17244 of SEQ ID NO: 26, the nucleic acid of residues 17278-17598 of SEQ ID NO: 26, the nucleic acid of residues 18880-19530 of SEQ ID NO: 26, the nucleic acid of residues 19795-19980 of SEQ ID NO: 26; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 27 and comprises a nucleic acid selected from the group consisting of: a) SEQ ID NO: 28; b) the nucleic acid of residues 103-1380 of SEQ ID NO: 28, the nucleic acid of residues 1450-2760 of SEQ ID NO: 28, the nucleic acid of residues 3583-4218 of SEQ ID NO: 28, the nucleic acid of residues 4468-4653 of SEQ ID NO: 28; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 29 and comprises a nucleic acid

selected from the group consisting of: a) SEQ ID NO: 30; b) the nucleic acid of residues 103-1380 of SEQ ID NO: 30, the nucleic acid of residues 1459-2754 of SEQ ID NO: 30, the nucleic acid of residues 3655-4293 of SEQ ID NO: 30, the nucleic acid of residues 4540-4725 of SEQ ID NO: 30, the nucleic acid of residues 4804-6081 of SEQ ID NO: 30, the nucleic acid of residues 6136-7419 of SEQ ID NO: 30, the nucleic acid of residues 7456-7776 of SEQ ID NO: 30, the nucleic acid of residues 8938-9588 of SEQ ID NO: 30, the nucleic acid of residues 9832-10017 of SEQ ID NO: 30, the nucleic acid of residues 10087-11364 of SEQ ID NO: 30, the nucleic acid of residues 11428-12711 of SEQ ID NO: 30, the nucleic acid of residues 12745-13065 of SEQ ID NO: 30, the nucleic acid of residues 14278-14928 of SEQ ID NO: 30, the nucleic acid of residues 15187-15372 of SEQ ID NO: 30; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 31 and comprises a nucleic acid selected from the group consisting of: a) SEQ ID NO: 32; b) the nucleic acid of residues 103-1380 of SEQ ID NO: 32, the nucleic acid of residues 1438-2742 of SEQ ID NO: 32, the nucleic acid of residues 2776-3096 of SEQ ID NO: 32, the nucleic acid of residues 4267-4917 of SEQ ID NO: 32, the nucleic acid of residues 5209-5394 of SEQ ID NO: 32, the nucleic acid of residues 5464-6741 of SEQ ID NO: 32, the nucleic acid of residues 6787-8070 of SEQ ID NO: 32, the nucleic acid of residues 8107-8427 of SEQ ID NO: 32, the nucleic acid of residues 9562-10212 of SEQ ID NO: 32, the nucleic acid of residues 10447-10632 of SEQ ID NO: 32, the nucleic acid of residues 10702-11979 of SEQ ID NO: 32, the nucleic acid of residues 12049-13326 of SEQ ID NO: 32, the nucleic acid of residues 13366-13686 of SEQ ID NO: 32, the nucleic acid of residues 14932-15582 of SEQ ID NO: 32, the nucleic acid of residues 15853-16038 of SEQ ID NO: 32; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 33 and comprises a nucleic acid

selected from the group consisting of: a) SEQ ID NO: 34; b) the nucleic acid of residues 103-1380 of SEQ ID NO: 34, the nucleic acid of residues 1441-2751 of SEQ ID NO: 34, the nucleic acid of residues 3613-4248 of SEQ ID NO: 34, the nucleic acid of residues 4498-4683 of SEQ ID NO: 34, the nucleic acid of residues 4753-6030 of SEQ ID NO: 34, the nucleic acid of residues 6199-7515 of SEQ ID NO: 34, the nucleic acid of residues 8356-8994 of SEQ ID NO: 34, the nucleic acid of residues 9247-9432 of SEQ ID NO: 34; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 35 and comprises a nucleic acid selected from the group consisting of: a) SED ID NO: 36; b) the nucleic acid of residues 118-1395 of SEQ ID NO: 36, the nucleic acid of residues 1507-2823 of SEQ ID NO: 36, the nucleic acid of residues 2860-3180 of SEQ ID NO: 36, the nucleic acid of residues 4366-5016 of SEQ ID NO: 36, the nucleic acid of residues 5251-5436 of SEQ ID NO: 36, the nucleic acid of residues 5503-6780 of SEQ ID NO: 36, the nucleic acid of residues 6841-8154 of SEQ ID NO: 36, the nucleic acid of residues 8191-8511 of SEQ ID NO: 36, the nucleic acid of residues 9562-10638 of SEQ ID NO: 36, the nucleic acid of residues 10651-11301 of SEQ ID NO: 36, the nucleic acid of residues 11536-11721 of SEQ ID NO: 36, the nucleic acid of residues 11794-13071 of SEQ ID NO: 36, the nucleic acid of residues 13117-14409 of SEQ ID NO: 36, the nucleic acid of residues 14443-14763 of SEQ ID NO: 36, the nucleic acid of residues 15898-16548 of SEQ ID NO: 36, the nucleic acid of residues 16789-16974 of SEQ ID NO: 36, the nucleic acid of residues 17056-18333 of SEQ ID NO: 36, the nucleic acid of residues 18391-19671 of SEQ ID NO: 36, the nucleic acid of residues 19714-20034 of SEQ ID NO: 36, the nucleic acid of residues 21184-21834 of SEQ ID NO: 36, the nucleic acid of residues 22087-22272 of SEQ ID NO: 36; c) a nucleic acid having at least 80% identity to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

In another embodiment the nucleic acid encodes one or more domains of the polyketide synthase of SEQ ID NO: 37 and comprises a nucleic acid

selected from the group consisting of: a) SEQ ID NO: 38; b) the nucleic acid of residues 100-1377 of SEQ ID NO: 38, the nucleic acid of residues 1504-2778 of SEQ ID NO: 38, the nucleic acid of residues 2812-3132 of SEQ ID NO: 38, the nucleic acid of residues 4258-4908 of SEQ ID NO: 38, the nucleic acid of residues 5143-5328 of SEQ ID NO: 38, the nucleic acid of residues 5395-6672 of SEQ ID NO: 38, the nucleic acid of residues 6739-8019 of SEQ ID NO: 38, the nucleic acid of residues 8056-8376 of SEQ ID NO: 38, the nucleic acid of residues 9607-10257 of SEQ ID NO: 38, the nucleic acid of residues 10537-10722 of SEQ ID NO: 38, the nucleic acid of residues 10945-11616 of SEQ ID NO: 38; c) a nucleic acid having at least 80% identical to a nucleic acid of a) or b); and d) a nucleic acid complementary to a nucleic acid of a), b) or c).

The invention also provides nucleic acids involved in the biosynthesis of a polyketide of Formula I other than those encoding a domain of the polyketide synthase system. In this embodiment, the invention provides an isolated, purified or enriched nucleic acid selected from the group consisting of: a) a nucleic acid of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 and 78; b) a nucleic acid encoding a polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 71, 73, 75 and 77; c) a nucleic acid having at least 75% identity to a nucleic acid of (a) or (b); and d) a nucleic acid complementary to a nucleic acid of (a), (b) or (c).

The invention further provides a nucleic acid that is hybridizable under stringent conditions to any one of the above nucleic acids and is substitutable for the nucleic acid to which it specifically hybridizes to direct the synthesis of a compound of Formula I. The invention further provides an isolated, purified or enriched nucleic acid comprising the sequence of at least two, preferably three, more preferably five, still more preferably 7 or more of the above nucleic acids.

The invention further provides an expression vector comprising any of the above nucleic acids. The invention further provides a host cell transformed with such an expression vector.

In a further aspect, the invention provides a gene cluster for production of a polyketide of Formula I. In one embodiment, the gene cluster may comprise at least ten, preferably twelve, more preferably fifteen, still more preferably twenty or more of the above nucleic acids. In a further embodiment, the gene cluster may include the nucleic acids of a cosmid selected from the cosmids deposited under IDAC accession nos. 250203-01, 250203-02, 250203-03, 250203-04, and 250203-05. In a further embodiment, the deposited cosmids are inserted into a prokaryotic host for expressing a product. The host may be *E. coli*, *Streptomyces lividans*, *Streptomyces griseofuscus*, *Streptomyces ambofaciens*, another species of *Actinomycetes*, or bacteria of the genus *Bacillus*, *Corynebacteria*, or *Thermoactinomyces*. In a further embodiment, the invention provides a nucleic acid which hybridizes under stringent hybridization conditions to the nucleic acids of the deposited cosmids and which encodes at least one protein involved in the biosynthesis of a polyene polyketide. In a further embodiment, the invention provides the isolated gene cluster from *Streptomyces aizunensis* encoding the biosynthetic pathway for the formation of compound 2(a), wherein said isolated gene cluster is the gene cluster formed by the deposited cosmids.

In another aspect, the invention relates to an isolated polypeptide for production of a polyketide of Formula I, and provides, in one embodiment, an amino acid sequence of a polyketide synthase domain of SEQ ID NO: 21, SEQ ID NO: 23, SEQ ID NO: 25, SEQ ID NO: 27, SEQ ID NO: 29, SEQ ID NO: 31, SEQ ID NO: 33, SEQ ID NO: 35 and SEQ ID NO: 37. The domain may be a β-ketoacyl synthase (KS) domain, an acyl carrier protein (ACP) domain, an acyl transferase (AT) domain, a ketoreductase (KR) domain, an enoyl reductase (ER) domain, a thioesterase (TE) domain or a dehydratase (DH) domain. In one embodiment, the domain is a KS domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of residues 141 to 566 of SEQ ID NO: 21, residues 1690 to 2118 of SEQ ID NO: 21, residues 3215 to 3640 of SEQ ID NO: 21, residues 5007 to 5438 of SEQ ID NO: 21, residues 6529 to 6954 of SEQ ID NO: 21, residues 37 to 462 of SEQ ID NO: 23, residues 1794 to 2215 of SEQ ID NO: 23, residues 36 to 461 of SEQ ID NO: 25, residues 1831 to 2256 of SEQ ID NO: 25, residues

3361 to 3786 of SEQ ID NO: 25, residues 4880 to 5304 of SEQ ID NO: 25, residues 35 to 460 of SEQ ID NO: 27, residues 35 to 460 of SEQ ID NO: 29, residues 1602 to 2027 of SEQ ID NO: 29, residues 3363 to 3788 of SEQ ID NO: 29, residues 35 to 460 of SEQ ID NO: 31, residues 1822 to 2247 of SEQ ID NO: 31, residues 3568 to 3993 of SEQ ID NO: 31, residues 35 to 460 of SEQ ID NO: 33, residues 1585 to 2010 of SEQ ID NO: 33, residues 40 to 465 of SEQ ID NO: 35, residues 1835 to 2260 of SEQ ID NO: 35, residues 3932 to 4357 of SEQ ID NO: 35, residues 5686 to 6111 of SEQ ID NO: 35, residues 34 to 459 of SEQ ID NO: 37, residues 1799 to 2224 of SEQ ID NO: 37; and amino acid sequence having at least 75% identity to any one of the above amino acid residues.

In another embodiment, the domain is an ACP domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of: residues 57 to 118 of SEQ ID NO: 21, residues 1603 to 1664 of SEQ ID NO: 21, residues 3130 to 3191 of SEQ ID NO: 21, residues 4911 to 4972 of SEQ ID NO: 21, residues 6444 to 6505 of SEQ ID NO: 21, residues 8002 to 8063 of SEQ ID NO: 21, residues 1706 to 1767 of SEQ ID NO: 23, residues 3258 to 3319 of SEQ ID NO: 23, residues 1736 to 1797 of SEQ ID NO: 25, residues 3278 to 3339 of SEQ ID NO: 25, residues 4795 to 4856 of SEQ ID NO: 25, residues 6599 to 6660 of SEQ ID NO: 25, residues 1490 to 1551 of SEQ ID NO: 27, residues 1514 to 1575 of SEQ ID NO: 29, residues 3278 to 3339 of SEQ ID NO: 29, residues 5060 to 5124 of SEQ ID NO: 29, residues 1737 to 1798 of SEQ ID NO: 31, residues 3483 to 3544 of SEQ ID NO: 31, residues 5285 to 5346 of SEQ ID NO: 31, residues 1500 to 1561 of SEQ ID NO: 33, residues 3083 to 3144 of SEQ ID NO: 33, residues 1751 to 1812 of SEQ ID NO: 35, residues 3846 to 3907 of SEQ ID NO: 35, residues 5597 to 5658 of SEQ ID NO: 35, residues 7363 to 7424 of SEQ ID NO: 35, residues 1715 to 1776 of SEQ ID NO: 37, residues 3513 to 3574 of SEQ ID NO: 37, and an amino acid sequence having at least 75% identity to any one of the above amino acid residues.

In another embodiment, the domain is a AT domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of: residues 597 to 1013 of SEQ ID NO: 21, residues 2135 to 2562 of SEQ ID

NO: 21, residues 3660 to 4089 of SEQ ID NO: 21, residues 5460 to 5883 of SEQ ID NO: 21, residues 6979 to 7402 of SEQ ID NO: 21, residues 493 to 919 of SEQ ID NO: 23, residues 2232 to 2659 of SEQ ID NO: 23, residues 483 to 907 of SEQ ID NO: 25, residues 2281 to 2714 of SEQ ID NO: 25, residues 3803 to 4225 of SEQ ID NO: 25, residues 5323 to 5748 of SEQ ID NO: 25, residues 484 to 920 of SEQ ID NO: 27, residues 487 to 918 of SEQ ID NO: 29, residues 2046 to 2473 of SEQ ID NO: 29, residues 3810 to 4237 of SEQ ID NO: 29, residues 480 to 914 of SEQ ID NO: 31, residues 2263 to 2690 of SEQ ID NO: 31, residues 4017 to 4442 of SEQ ID NO: 31, residues 481 to 917 of SEQ ID NO: 33, residues 2067 to 2505 of SEQ ID NO: 33, residues 503 to 941 of SEQ ID NO: 35, residues 2281 to 2718 of SEQ ID NO: 35, residues 4373 to 4803 of SEQ ID NO: 35, residues 6131 to 6557 of SEQ ID NO: 35, residues 502 to 926 of SEQ ID NO: 37, residues 2247 to 2673 of SEQ ID NO: 37; and an amino acid sequence having at least 75% identity to any one of the above amino acid residues.

In another embodiment, the domain is a KR domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of: residues 1304 to 1517 of SEQ ID NO: 21, residues 2833 to 3045 of SEQ ID NO: 21, residues 4612 to 4829 of SEQ ID NO: 21, residues 6147 to 6360 of SEQ ID NO: 21, residues 7703 to 7918 of SEQ ID NO: 21, residues 1411 to 1627 of SEQ ID NO: 23, residues 2960 to 3173 of SEQ ID NO: 23, residues 1439 to 1655 of SEQ ID NO: 25, residues 2981 to 3194 of SEQ ID NO: 25, residues 4494 to 4706 of SEQ ID NO: 25, residues 6294 to 6510 of SEQ ID NO: 25, residues 1195 to 1406 of SEQ ID NO: 27, residues 1219 to 1431 of SEQ ID NO: 29, residues 2980 to 3196 of SEQ ID NO: 29, residues 4760 to 4976 of SEQ ID NO: 29, residues 1423 to 1639 of SEQ ID NO: 31, residues 3188 to 3404 of SEQ ID NO: 31, residues 4978 to 5194 of SEQ ID NO: 31, residues 1205 to 1416 of SEQ ID NO: 33, residues 2786 to 2998 of SEQ ID NO: 33, residues 1456 to 1672 of SEQ ID NO: 35, residues 3551 to 3767 of SEQ ID NO: 35, residues 5300 to 5516 of SEQ ID NO: 35, residues 7062 to 7288 of SEQ ID NO: 35, residues 1420 to 1636 of SEQ ID NO: 37, residues 3203 to 3419 of SEQ ID NO: 37; and an amino acid sequence having at least 75% identity to any one of the above amino acid residues.

In another embodiment, the domain is a DH domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of: residues 4102 to 4208 of SEQ ID NO: 21, residues 932 to 1038 of SEQ ID NO: 23, residues 919 to 1027 of SEQ ID NO: 25, residues 5761 to 5866 of SEQ ID NO: 25, residues 2486 to 2592 of SEQ ID NO: 29, residues 4249-4355 of SEQ ID NO: 29 residues 926 to 1032 of SEQ ID NO: 31, residues 2703 to 2809 of SEQ ID NO: 31, residues 4456 to 4562 of SEQ ID NO: 31, residues 954 to 1060 of SEQ ID NO: 35, residues 2731 to 2837 of SEQ ID NO: 35, residues 4815 to 4921 of SEQ ID NO: 35, residues 6572 to 6678 of SEQ ID NO: 35, residues 938 to 1044 of SEQ ID NO: 37; residues 2686 to 2792 of SEQ ID NO: 37; and an amino acid sequence having at least 75% identity to any one of the above amino acid residues.

In another embodiment, the domain is an ER domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of: residues 3188 to 3546 of SEQ ID NO: 35 and any amino acid sequence having at least 75% identity to residues 3188 to 3546 of SEQ ID NO: 35.

In another embodiment, the domain is an TE domain and the amino acid comprises a sequence selected from the group consisting of the amino acid of: residues 3649 to 3872 of SEQ ID NO: 37, and any amino acid sequence having at least 75% identity to residues 3649 to 3872 of SEQ ID NO: 37.

In another embodiment, the invention provides a polypeptide involved in the biosynthesis of a polyketide of Formula I other than a polypeptide encoding a domain of the polyketide synthase system of the invention. In this embodiment, the invention provides an isolated polypeptide for the production of a polyketide of Formula I selected from the group consisting of: a) SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75 and 77; and b) a polypeptide which is at least 75% identical to SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75 and 77.

In another aspect, the invention provides a method of making a polypeptide having a sequence selected from the group consisting of SEQ ID

NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75 and 77 comprising the steps of: (a) introducing a nucleic acid encoding said polypeptide, said nucleic acid being operably linked to a promoter, into a bacterial host cell; and (b) culturing the transformed host cell under conditions which result in the expression of the polypeptide.

In another aspect the invention is drawn to a method for increasing the yield of the polyketides of the invention using the deposited cosmids of the nucleic acids described above, said method comprising the steps of transforming a prokaryotic host with cosmids or nucleic acids and culturing the transformed prokaryotic host under conditions which result in the expression of the polyketide.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Diagram of the biosynthetic locus for compound 2(a) from *Streptomyces aizunensis*. Also indicated are the positions of cosmids depositedunder IIDAC accession numbers 250203-01, 250203-02, 250203-03, 250203-04 and 250203-05, which span the locus of compound 2(a).

Figure 2a-d: Multiple amino acid alignment comparing the 26 KS domains present in the polyketide synthase (PKS) for compound 2(a) (ORFs 10 to 18). The boundaries and key residues (highlighted in black) of the KS domains were chosen as described by Kakavas *et al.*, *J. Bacteriol.*179, 7515-7522 (1997).

Figure 3a-d: Multiple amino acid alignment comparing the 26 AT domains present in the compound 2(a) PKS (ORFs 10 to 18). The boundaries and key residues (highlighted in black) of the AT domains were chosen as described by Kakavas *et al.*, *supra*.

Figure 4: Multiple amino acid alignment comparing the 15 DH domains present in the compound 2(a) PKS (ORFs 10, 11, 12, 14, 15, 17 and 18). The boundaries and key residues (highlighted in black) of the DH domains were chosen as described by Kakavas *et al. supra*. The inactive DH domains are highlighted.

Figure 5: Amino acid alignment comparing the ER domain present in the compound 2(a) PKS (ORF 17) with the ER domains from modules 5 and

15 in the nystatin biosynthetic locus as described by Brautaset *et al.*, *Chem. Biol.*, 7, 395-403 (2000). The boundaries and key residues (highlighted in black) of the ER domain were chosen as described by Kakavas *et al. supra*.

Figure 6a and 6b: Multiple amino acid alignment comparing the 26 KR domains present in the compound 2(a) PKS (ORFs 10 to 18). The boundaries and key residues (highlighted in black) of the KR domains were chosen as described by Kakavas *et al. supra*, and Fisher *et al. Structure Fold Des.* 8, 339-347 (2000). The inactive KR domain found in ORF 13/module 12 is highlighted.

Figure 7: Multiple amino acid alignment comparing the 27 ACP domains present in the compound 2(a) PKS (ORFs 10 to 18). The boundaries and key serine residues (highlighted in black) of the ACP domains were chosen as described by Kakavas *et al. supra*.

Figure 8: Amino acid alignment comparing the TE domain present in the compound 2(a) PKS (ORF 18) with the TE domain from module 7 in the nystatin biosynthetic locus as described by Brautaset *et al. supra*. The boundaries and key residues (highlighted in black) of the ER domain were chosen as described by Kakavas *et al. supra*.

In each of the clustal alignments (Figs 2 to 8) a line below the alignment is used to mark strongly conserved positions. In addition, three characters, namely * (asterisk), : (colon) and . (period) are used, wherein "*" indicates positions which have a single, fully conserved residue; ":" indicates that one of the following strong groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, and FYW; and "." indicates that one of the following weaker groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, and HFY.

Figure 9: Phylogenetic analysis of the 26 AT domains present in the compound 2(a) PKS (ORFs 10 to 18) along with a malonyl-specific and a methylmalonyl-specific AT domain present in modules 3 and 11 respectively of the nystatin PKS system as described by Brautaset *et al. supra*.

Figure 10a to 10c: biosynthetic pathway for compound 2(a) polyketide core structure.

Figure 11a and 11b: biosynthetic pathways for compound 2(a) aminohydroxy-cyclopentenone (a) and deoxysugar (b) components.

Figures 12a to 12f: outline of strategies for the genetic modification of locus for compound 2(a) providing for variants that functionally modify compound 2(a).

Figure 13: shows the data for the compound of compound 2(a) obtained by electrospray mass spectrometry.

Figure 14: shows the data for the compound of compound 2(a) obtained by UV λ_{max}.

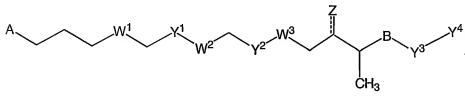
Figure 15: shows the data obtained for the compound of compound 2(a) by NMR at 500 MHz dissolved in d_3-MeOH including proton 15 A, carbon 15 B, and multidimensional pulse sequences gDQCOSY, gHSQC, gHMBC, and TOCSY 15 C, 15D, 15E and 15F, respectively.

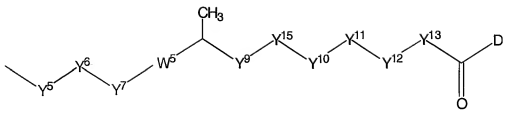
Figure 16: is a plot of the data from a study to evaluate the antifungal activity of compound 2(a) against *Candida albicans* in a mouse model as described in Example 5. Figure 16 depicts the percent survival versus days post-inoculation with compound 2(a) (3 mg/kg), compound 2(a) (1 mg/kg), Fungizone (0.25 mg/kg) and Fungizone (0.50 mg/kg).

Figure 17: proton-NMR (Figure 17A) and carbon-13 NMR (Figure 17B) spectral assignments for Compound 2(a) as discussed in Example 3.

DETAILED DESCRIPTION OF THE INVENTION

The present invention encompasses compounds of Formula I, and pharmaceutically acceptable salts thereof:

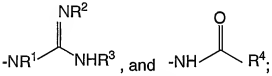




Formula I

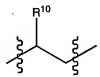
wherein,

A is selected from the group consisting of $-NR^1R^2$, $-N=CR^1R^2$,



$-NR^1$, NR^2, NHR^3, and $-NH$, O, R^4;

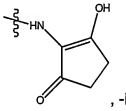
R^1, R^2, R^3 and R^4 are each independently selected from the group consisting of H, C_{1-6} alkyl, C_{2-6} alkenyl, C_{3-6} cycloalkyl, C_{2-6} heterocycloalkyl, aryl, heteroaryl and amino acid, wherein said alkyl, alkenyl, aryl and heteroaryl are optionally substituted with a group selected from halogen, OH, NO_2, NH_2 or aryl, said aryl being optionally further substituted with one or more groups independently selected from halogen, OH, NO_2 or NH_2;

B is selected from ethene-1,2-diyl or  ;

wherein R^{10} is oxo or OR^{11};

wherein R^{11} is H or a heterocycloalkyl, the heterocycloalkyl being optionally substituted with 1-4 substituents selected from OX, C_{1-3} alkyl and $-O-C(O)R^1$, wherein X is H or, when there are at least two neighboring substituent groups that are OX, then the X can be a bond such that the two neighboring oxygen groups form a five-membered acetal ring of the formula:

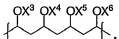

; wherein R^5 and R^6 are each independently selected from the group consisting of H, C_{1-6} alkyl, and C_{2-7} alkenyl;

D is selected from:  , $-NR^{12a}R^{12a}$, and OR^{12}, wherein

R^{12} is selected from H, C_{1-6} alkyl optionally substituted with 1 to 2 phenyl groups, wherein the phenyl group is optionally substituted with C_{1-6} alkyl and halo;

R^{12a} and R^{12a} are each indepedently selected from H, C_{1-6} alkyl, C_{2-6} alkenyl, C_{3-6} cycloalkyl, C_{2-6} heterocycloalkyl, aryl, heteroaryl and amino acid, wherein said alkyl, alkenyl, aryl and heteroaryl are optionally substituted with a group selected from halogen, OH, NO_2, NH_2 or aryl, said aryl being optionally further substituted with one or more groups independently selected from halogen, OH, NO_2 or NH_2;

W^1 is  ;

W^2 is  ;

W^3 is  ;

W^5 is  ;

X^1, X^2, X^3, X^4, X^5, X^6, X^7, X^8, X^9, X^{12} and X^{13} are each independently selected from H, -C(O)-R^7 and a bond such that when any of two neighboring X^1, X^2, X^3, X^4, X^5, X^6, X^7, X^8, X^9, X^{12} and X^{13} is a bond then the two neighboring oxygen atoms and their attached carbon atoms together form a six-membered acetal ring of the formula:

R^5, R^6 and R^7 are each independently selected from H, C_{1-6} alkyl, C_{2-7} alkenyl;

Y^1, Y^2, Y^3, Y^4, Y^5, Y^6, Y^7, Y^9, Y^{10}, Y^{11}, Y^{12}, Y^{13} and Y^{15} are each independently selected from the group consisting of ethene-1,2-diyl, ethane-1,2-diyl and  , wherein said ethene-1,2-diyl and ethane-1,2-diyl groups are optionally substituted with a methyl group;

Z is selected from OH, NHR^8,  and when the dotted line is a bond then Z is oxo, or NR^9;

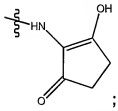
R^8 is selected from H, C_{1-6} alkyl, C_{2-6} alkenyl;

R^9 is C_{1-6} alkyl optionally substituted with aryl.

In a first embodiment the invention provides compounds of Formula I wherein Z is oxo; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Within this first embodiment Z is oxo, A is -NR^1R^2; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Further within this embodiment Z is oxo, A is - NR^1R^2; and D is

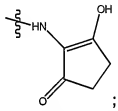


; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Within the first embodiment the invention provides compounds of Formula I wherein Z is oxo and A is

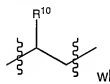
; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Further within this embodiment Z is oxo and A is -NH-C(=O)-R^4 and D is



; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

In a second embodiment the invention provides compounds of Formula I wherein B is



wherein R^{10} is oxo or OR^{11}; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Within this second embodiment R^{10} is OR^{11}, wherein R^{11} is a heterocycloalkyl, the heterocycloalkyl being optionally substituted with 1-4 substituents selected from OX, C_{1-3} alkyl and -O-C(O)R^1, wherein X is H or,

when there are at least two neighboring substituent groups that are OX, then the X can be a bond such that the two neighboring oxygen groups form a five-membered acetal ring of the formula:

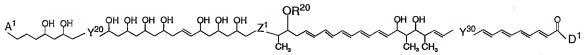
Within this embodiment R^{11} is a heterocycloalkyl, the heterocycloalkyl being optionally substituted with 1-4 substituents selected from OX, C_{1-3} alkyl and $-O-C(O)R^1$, wherein X is H or, when there are at least two neighboring substituent groups that are OX, then the X can be a bond such that the two neighboring oxygen groups form a five- membered acetal ring of the formula:

 , and A is $-NR^1R^2$; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Further within this embodiment the invention provides compounds of Formula I, wherein R^{11} is a heterocycloalkyl, the heterocycloalkyl being optionally substituted with 1-4 substituents selected from OX, C_{1-3} alkyl and $-O-C(O)R^1$, wherein X is H or, when there are at least two neighboring substituent groups that are OX, then the X can be a bond such that the two neighboring oxygen groups form a five-membered acetal ring of the formula:

 , A is $-NR^1R^2$ and Z is oxo; and all other groups are as previously defined; or a pharmaceutically acceptable salt thereof.

Preferred compounds of the invention comprise compounds of Formula II:



Formula II

wherein A^1 is $-NH_2$, $-N=CH-R^{13}$, amino acid or $-NH-R^{14}$, wherein R^{13} is hydrogen or phenyl and R^{14} is selected from the group consisting of isopropyl, 1-(4-nitrophenyl)methyl, cyclohexyl, and wherein said amino acid is attached via its nitrogen atom;

and

wherein R^{15} is selected from the group consisting of methyl, isopropyl, phenyl, 4-nitrophenyl, 1-aminoethyl, 1-amino-1-(4-hydroxyphenyl)methyl, 1-amino-2-(4-hydroxyphenyl)ethyl, 1-amino-2-methylpropyl, 2-pyrrolidinyl and1-amino-2-hydroxyethyl;

Y^{20} is selected from the group consisting of ethene-1,2-diyl and

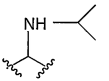
;

Z^1 is selected from the group consisting of:

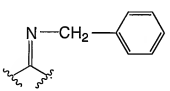
,

,

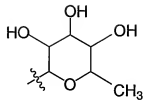
,



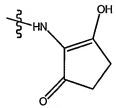
and



R^{20} is selected from the group consisting of hydrogen and



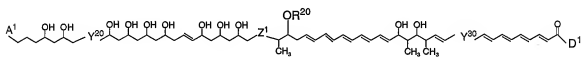
Y^{30} is ethene-1,2-diyl or ethane-1,2-diyl; and
D^1 is hydroxy, methoxy or



and pharmaceutically acceptable salts thereof.

The present invention includes pharmaceutical compositions of the compounds of Formula II, said compositions comprising a therapeutically effective amount of the compound of Formula II or a pharmaceutically acceptable salt thereof, and a pharmaceutically acceptable carrier.

Particularly preferred compounds of the present invention include those of Formula II



Formula II

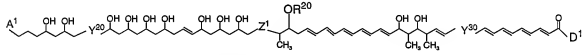
wherein A^1 is amino ($-NH_2$), and Y^{20}, Z^1, R^{20}, Y^{30} and D^1 are as defined in Table A below.

Table A. Compounds of Formula II wherein A^1 is NH_2

Compound	Y^{20}	Z^1	R^{20}	Y^{30}	D^1
2(a)	ethene-1,2-diyl		3,4,5-trihydroxy-6-methyl-tetrahydro-pyran-2-yl	ethane-1,2-diyl	

2(b)		"	"	"	"
2(c)	ethene-1,2-diyl		"	"	"
2(d)	"		"	"	"
2(e)	"		"	"	"
2(f)	"		"	"	"
2(g)	"		"	"	hydroxy
2(h)	"	"	"	"	methoxy
2(i)	"	"	hydrogen	"	
2(j)	"	"	"	"	hydroxy
2(k)	"	"	3,4,5-trihydroxy-6-methyl-tetrahydro-pyran-2-yl	ethene-1,2-diyl	
2(l)	"		"	"	"

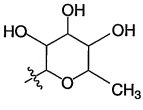
Additional preferred compounds of the invention include compounds of Formula II



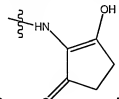
Formula II

as set forth in Tables B and C below,

wherein Y^{20} is ethene-1,2-diyl;

Z^1 is  ; R^{20} is  ;

Y^{30} is ethane-1,2-diyl; and

D^1 is  ; and

wherein A^1 is –N=CH-R^{13} (Table B); -NH-R^{14} (Table C).

Table B. Compounds of Formula II wherein A^1 is –N=CH-R^{13} and Y^{20}, Z^1, R^{20}, Y^{30} and D^1 are as defined above.

Compound	R^{13}
2(m)	CH_3
2(n)	phenyl

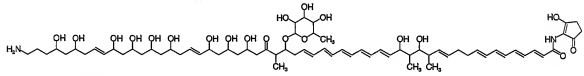
Table C. Compounds of Formula II wherein A^1 is -NH-R^{14} and Y^{20}, Z^1, R^{20}, Y^{30} and D^1 are as defined above.

Compound	R^{14}	R^{15}
2(o)		NA*
2(p)	isopropyl	NA
2(q)	1-(4-nitrophenyl)methyl	NA
2(r)	cyclohexyl	NA
2(s)		CH_3
2(t)		isopropyl
2(u)		phenyl
2(v)		4-nitrophenyl
2(w)		1-aminoethyl
2(x)		1-amino-1-(4-hydroxyphenyl)methyl
2(y)		1-amino-2-(4-hydroxyphenyl)ethyl

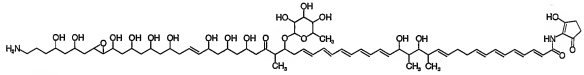
		hydroxyphenyl)ethyl
2(z)		1-amino-2-methylpropyl
2(aa)		2-pyrrolidinyl
2(ab)		1-amino-2-hydroxyethyl

*NA = not applicable

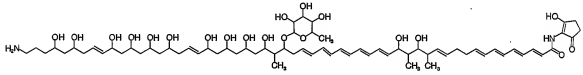
The compounds of Tables A, B and C are shown below.



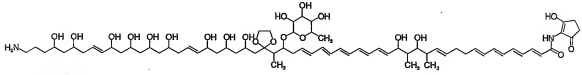
Compound 2(a)



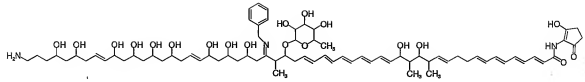
Compound 2(b)



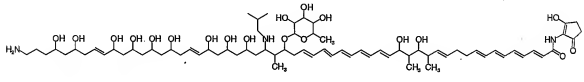
Compound 2(c)



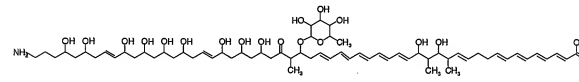
Compound 2(d)



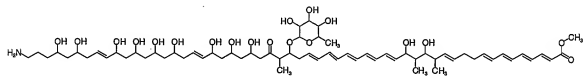
Compound 2(e)



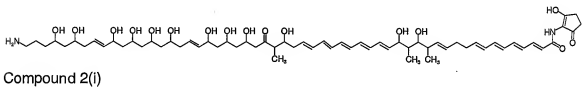
Compound 2(f)



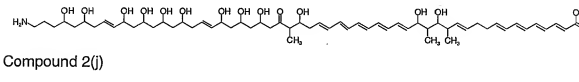
Compound 2(g)



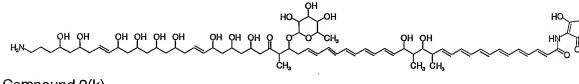
Compound 2(h)



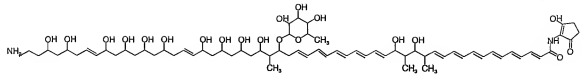
Compound 2(i)



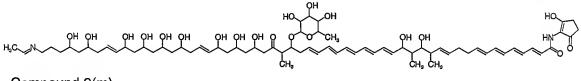
Compound 2(j)



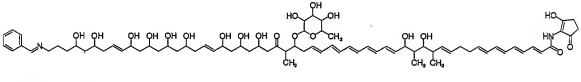
Compound 2(k)



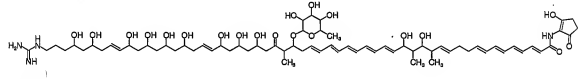
Compound 2(l)



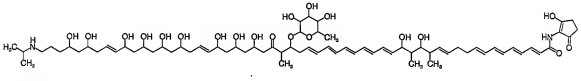
Compound 2(m)



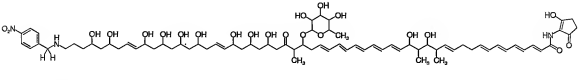
Compound 2(n)



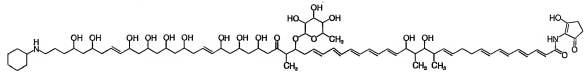
Compound 2(o)



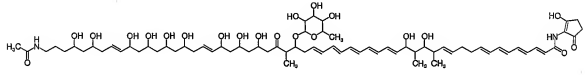
Compound 2(p)



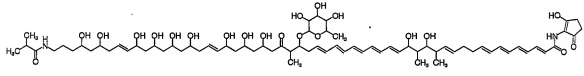
Compound 2(q)



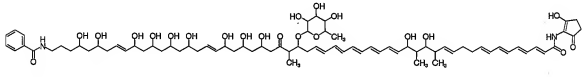
Compound 2(r)



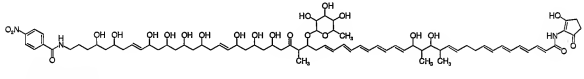
Compound 2(s)



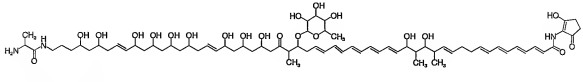
Compound 2(t)



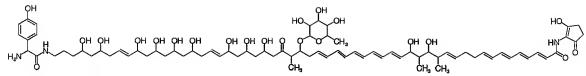
Compound 2(u)



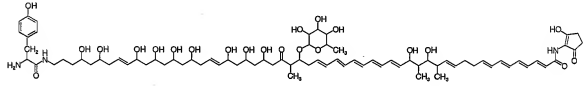
Compound 2(v)



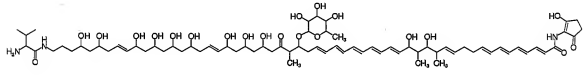
Compound 2(w)



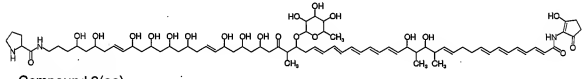
Compound 2(x)



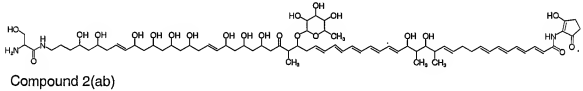
Compound 2(y)



Compound 2(z)



Compound 2(aa)



Compound 2(ab)

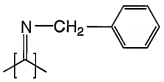
The following bivalent moieties are referred to herein by the nomenclature as indicated below:

1-oxo-methylene-1,1-diyl

1-hydroxymethylene-1,1-diyl

1,3-dioxacyclopentane-2,2-diyl

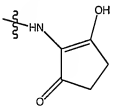
(2-propylamino)methylene-1,1-diyl



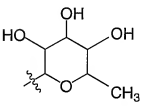
1-benzyliminomethylene-1,1-diyl

oxirane-2,3-diyl.

The following monovalent moieties are referred to herein by the nomenclature as indicated:



(2-hydroxy-5-oxo-cyclopent-1-enyl)-amino



3,4,5-trihydroxy-6-methyl-tetrahydropyran-2-yl.

The terms "polyketide" or "polyene polyketide" refer to a class of polyketide compounds defined by Formula I or II. A preferred polyketide of

the invention is the compound 2a, having the systematic name 56-Amino-15,17,33,35,37,41,43,45,47,51,53-undecahydroxy-14,16,30-trimethyl-31-oxo-29-(3,4,5-trihydroxy-6-methyl-tetrahydro-pyran-2-yloxy)-hexapentaconta-2,4,6,8,12,18,20,22,24,26,38,48-dodecaenoic acid (2-hydroxy-5-oxo-cyclopent-1-enyl)-amide. The term further includes compounds of this class that can be used as intermediates in chemical synthesis.

The terms "producer of compounds of Formula I" and "compounds of Formula I -producing organism" refer to a microorganism that carries genetic information necessary to produce a compound of Formula I, whether or not the organism is known to produce a compound of Formula I. The terms "producer of compounds of Formula II" and "compound of Formula II-producing organism" refer to a microorganism that carries genetic information necessary to produce a compound of Formula II, whether or not the organism is known to produce a compound of Formula II. The terms "producer of Compound 2(a)" and "Compound 2(a)-producing organism" refer to a microorganism that carries genetic information necessary to produce Compound 2(a), whether or not the organism is known to produce Compound 2(a). The term "polyketide producer" refer to a microorganism that carries genetic information necessary to produce a polyketide of Formula I or II. The terms apply equally to organisms in which the genetic information to produce the compound of Formula I or II or Compound 2(a) is found in the organism as it exists in its natural environment, and to organisms in which the genetic information is introduced by recombinant techniques. For the sake of particularity, specific organisms contemplated herein include organisms of the family *Micromonosporaceae*, of which preferred genera include *Micromonospora, Actinoplanes* and *Dactylosporangium*; the family *Streptomycetaceae*, of which preferred genera include *Streptomyces* and *Kitasatospora*; the family *Pseudonocardiaceae*, of which preferred genera are *Amycolatopsis* and *Saccharopolyspora;* and the family *Actinosynnemataceae*, of which preferred genera include *Saccharothrix* and *Actinosynnema*; however the terms are intended to encompass all organisms containing genetic information necessary to produce a compound of Formula I or II or Compound 2(a). Preferred producers of a compound of formula I or II or Compound 2(a)

include *Streptomyces aizunensis* (NRRL B-11277) and any mutant or improved strain of *Streptomyces aizunensis*, including strain [C03]023 (IDAC accession no. 070803-01) and strain [C03U03]023 (IDAC accession no. 231203-02).

The term "isolated" means that the material is removed from its original environment, *e.g.* the natural environment if it is naturally-occurring. For example, a naturally occurring polynucleotide or polypeptide present in a living organism is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Individual nucleic acids obtained from a library have been conventionally purified to electrophoretic homogeneity. The purified nucleic acids of the present invention have been purified from the remainder of the genomic DNA in the organism by at least 10^4 to 10^6 fold. However, the term "purified" also includes nucleic acids which have been purified from the remainder of the genomic DNA or from other sequences in a library or other environment by at least one order of magnitude, preferably two or three orders of magnitude, and more preferably four or five orders of magnitude.

"Recombinant" means that the nucleic acid is present in the cell with "backbone" nucleic acid, wherein the nucleic acid is not present with "backbone" nucleic acid in its natural environment. "Recombinant" can also be defined to mean that the nucleic acid is adjacent to "backbone" nucleic acid to which it is not adjacent in its natural environment. "Enriched" nucleic acids represent 5% or more of the number of nucleic acid inserts in a population of nucleic acid backbone molecules. "Backbone" molecules include nucleic acids such as expression vectors, self-replicating nucleic acids, viruses, integrating nucleic acids, and other vectors or nucleic acids used to maintain or manipulate a nucleic acid of interest. Preferably, the enriched nucleic acids

represent 15% or more, more preferably 50% or more, and most preferably 90% or more, of the number of nucleic acid inserts in the population of recombinant backbone molecules.

"Recombinant" polypeptides or proteins refer to polypeptides or proteins produced by recombinant DNA techniques, *i.e.* produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide or protein. "Synthetic" polypeptides or proteins are those prepared by chemical synthesis.

The term "gene" means the segment of DNA involved in producing a polypeptide chain; it includes regions preceding and following the coding region (leader and trailer) as well as, where applicable, intervening regions (introns) between individual coding segments (exons).

The terms "gene locus, "gene cluster," and "biosynthetic locus" refer to a group of genes or variants thereof involved in the biosynthesis of the polyketide of Formula 2a. Genetic modification of gene locus, gene cluster or biosynthetic locus refers to any genetic recombinant techniques known in the art including mutagenesis, inactivation, or replacement of nucleic acids that can be applied to generate variants of the compounds of Formula 2a. Genetic modification of gene locus, gene cluster or biosynthetic locus refers to any genetic recombinant techniques known in the art including mutagenesis, inactivation, or replacement of nucleic acids that can be applied to generate genetic variants of compounds of Formula I.

A DNA or nucleotide "coding sequence" or "sequence encoding" a particular polypeptide or protein, is a DNA sequence which is transcribed and translated into a polypeptide or protein when placed under the control of appropriate regulatory sequences.

"Oligonucleotide" refers to a nucleic acid, generally of at least 10, preferably 15 and more preferably at least 20 nucleotides, preferably no more than 100 nucleotides, that are hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule encoding a gene, mRNA, cDNA or other nucleic acid of interest.

A promoter sequence is "operably linked to" a coding sequence recognized by RNA polymerase which initiates transcription at the promoter and transcribes the coding sequence into mRNA.

"Digestion" of DNA refers to enzymatic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinary skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion, gel electrophoresis may be performed to isolate the desired fragment.

As used herein and as known in the art, the term "identity" is the relationship between two or more polynucleotide sequences, as determined by comparing the sequences. Identity also means the degree of sequence relatedness between polynucleotide sequences, as determined by the match between strings of such sequences. Identity can be readily calculated (see, e.g., Computation Molecular Biology, Lesk, A.M., eds., Oxford University Press, New York (1998), and Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York (1993), both of which are incorporated by reference herein). While there exist a number of methods to measure identity between two polynucleotide sequences, the term is well known to skilled artisans (see, e.g., Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press (1987); and Sequence Analysis Primer, Gribskov., M. and Devereux, J., eds., M. Stockton Press, New York (1991)). Methods commonly employed to determine identity between sequences include, for example, those disclosed in Carillo, H., and Lipman, D., SIAM J. Applied Math. (1988) 48:1073. "Substantially identical," as used herein, means there is a very high degree of homology (preferably 100% sequence

identity) between subject polynucleotide sequences. However, polynucleotides having greater than 90%, or 95% sequence identity may be used in the present invention, and thus sequence variations that might be expected due to genetic mutation, strain polymorphism, or evolutionary divergence can be tolerated.

The biosynthetic locus for the production of the Compound 2(a) spans approximately 176,000 base pairs of DNA and encodes 38 proteins. More than 10 kilobases of DNA sequence were analyzed on each side of the locus and these regions were found to contain primary metabolic genes.
The order and relative position of the 38 open reading frames representing the proteins of the biosynthetic locus for Compound 2(a) are provided in Figure 1. Referring to Figure 1, the genes involved in the biosynthesis of Compound 2(a) are contained within two contiguous nucleotide sequences (SEQ ID NOS: 1 and 18). The contiguous nucleotide sequences are arranged such that, as found within the compound 2(a) biosynthetic locus, the 3' end of the 11740 base pairs of DNA of contig 1 (SEQ ID NO: 1) is found adjacent to the 5' end of the 164,051 base pairs of DNA of contig 2 (SEQ ID NO: 18).

The nucleotide sequence and polypeptide sequences relating to the locus of compound 2(a) are provided in the sequence listing filed together with and forming part of this application. SEQ ID NO: 1 is the 11740 contiguous base pairs of contig 1 comprising eight open reading frames, namely ORF 1 to ORF 8 listed in SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15 and 17 respectively. The gene product of ORF 1 (SEQ ID NO: 2) is the 719 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 3 which is drawn from residues 418 to 2577 (sense strand) of contig 1 (SEQ ID NO: 1). The gene product of ORF 2 (SEQ ID NO: 4) is the 253 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 5 which is drawn from residues 3006 to 3767 (sense strand) of contig 1 (SEQ ID NO: 1). The gene product of ORF 3 (SEQ ID NO: 6) is the 956 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 7 which is drawn from residues 4016 to 6886 (sense strand) of contig 1 (SEQ ID NO: 1). The gene product of ORF 4 (SEQ ID NO: 8) is the 201 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 9 which is drawn from residues 7581 to 6976 (antisense strand) of contig 1 (SEQ ID

NO: 1). The gene product of ORF 5 (SEQ ID NO: 10) is the 416 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 11 which is drawn from residues 8848 to 7598 (antisense strand) of contig 1 (SEQ ID NO: 1). The gene product of ORF 6 (SEQ ID NO: 12) is the 186 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 13 which is drawn from residues 9053 to 9613 (sense strand) of contig 1 (SEQ ID NO: 1). The gene product of ORF 7 (SEQ ID NO: 14) is the 163 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 15 which is drawn from residues 9682 to 10173 (sense strand) of contig 1 (SEQ ID NO: 1). The gene product of ORF 8 (SEQ ID NO: 16) is the 514 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 17 which is drawn from residues 10170 to 11714 (sense strand) of contig 1 (SEQ ID NO: 1).

SEQ ID NO: 18 is the 164,051 contiguous base pairs of contig 2 comprising 30 ORFs, namely ORF 9 to ORF 38 listed in SEQ ID NOS: 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 and 78 respectively. The gene product of ORF 9 (SEQ ID NO: 19) is the 367 amino acids deduced from the nucleic acids sequence of SEQ ID NO: 20 which is drawn from residues 1109 to 6 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 10 (SEQ ID NO: 21) is the 8147 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 22 which is drawn from residues 1375 to 25818 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 11 (SEQ ID NO: 23) is the 3428 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 24 which is drawn from residues 25902 to 36188 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 12 (SEQ ID NO: 25) is the 6751 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 26 which is drawn from residues 36213 to 56468 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 13 (SEQ ID NO: 27) is the 1657 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 28 which is drawn from residues 56600 to 61573 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 14 (SEQ ID NO: 29) is the 5207 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 30 which is drawn from residues 61852 to 77475

(sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 15 (SEQ ID NO: 31) is the 5432 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 32 which is drawn from residues 77606 to 93904 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 16 (SEQ ID NO: 33) is the 3227 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 34 which is drawn from residues 94057 to 103740 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 17 (SEQ ID NO: 35) is the 7510 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 36 which is drawn from residues 103789 to 126321 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 18 (SEQ ID NO: 37) is the 3872 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 38 which is drawn from residues 126389 to 138007 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 19 (SEQ ID NO: 39) is the 338 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 40 which is drawn from residues 139079 to 138063 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 20 (SEQ ID NO: 41) is the 283 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 42 which is drawn from residues 140117 to 139266 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 21 (SEQ ID NO: 43) is the 329 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 44 which is drawn from residues 141103 to 140114 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 22 (SEQ ID NO: 45) is the 317 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 46 which is drawn from residues 141483 to 142436 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 23 (SEQ ID NO: 47) is the 204 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 48 which is drawn from residues 142440 to 143054 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 24 (SEQ ID NO: 49) is the 328 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 50 which is drawn from residues 143133 to 144119 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 25 (SEQ ID NO: 51) is the 328 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 52 which is drawn from residues 144116 to 145102

(sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 26 (SEQ ID NO: 53) is the 214 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 54 which is drawn from residues 145099 to 145743 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 27 (SEQ ID NO: 55) is the 470 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 56 which is drawn from residues 145818 to 147230 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 28 (SEQ ID NO: 57) is the 553 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 58 which is drawn from residues 148967 to 147306 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 29 (SEQ ID NO: 59) is the 231 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 60 which is drawn from residues 149871 to 149176 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 30 (SEQ ID NO: 61) is the 306 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 62 which is drawn from residues 150788 to 149868 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 31 (SEQ ID NO: 63) is the 998 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 64 which is drawn from residues 153765 to 150769 (antisense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 32 (SEQ ID NO: 65) is the 518 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 66 which is drawn from residues 154485 to 156041 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 33 (SEQ ID NO: 67) is the 329 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 68 which is drawn from residues 156075 to 157064 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 34 (SEQ ID NO: 69) is the 521 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 70 which is drawn from residues 157308 to 158873 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 35 (SEQ ID NO: 71) is the 410 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 72 which is drawn from residues 158970 to 160202 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 36 (SEQ ID NO: 73) is the 506 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 74 which is drawn from residues 160199 to 161719

(sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 37 (SEQ ID NO: 75) is the 217 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 76 which is drawn from residues 161924 to 162577 (sense strand) of contig 2 (SEQ ID NO: 18). The gene product of ORF 38 (SEQ ID NO: 77) is the 442 amino acids deduced from the nucleic acid sequence of SEQ ID NO: 78 which is drawn from residues 162723 to 164051 (sense strand) of contig 2 (SEQ ID NO: 18).

Some open reading frames listed herein initiate with non-standard initiation codons (*e.g.* GTG – Valine or CTG - Leucine) rather than the standard initiation codon ATG, namely ORFs 3, 5, 6, 9, 11, 13, 21, 22, 23, 24, 27, 34, 36 and 37 (SEQ ID NOS: 7, 11, 13, 20, 24, 28, 44, 46, 48, 50, 56, 70, 74 and 76). All ORFs are listed with the appropriate M, V or L amino acids at the amino-terminal position to indicate the specificity of the first codon of the ORF. It is expected, however, that in all cases the biosynthesized protein will contain a methionine residue, and more specifically a formylmethionine residue, at the amino terminal position, in keeping with the widely accepted principle that protein synthesis in bacteria initiates with methionine (formylmethionine) even when the encoding gene specifies a non-standard initiation codon (e.g. Stryer, Biochemistry 3rd edition, 1998, W.H. Freeman and Co., New York, pp. 752-754).

Five *E. coli* DH10B deposits, each harbouring a cosmid clone of a partial biosynthetic locus for compound 2(a) from *Streptomyces aizunensis* (NRRL B-11277) and together spanning the full locus were deposited with the International Depositary Authority of Canada, Bureau of Microbiology, Health Canada, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2 on February 25, 2003 and were assigned deposit accession numbers IDAC 250203-01, IDAC 250203-02, IDAC 250203-03, IDAC 250203-04 and IDAC 250203-05 respectively. The sequence of the polynucleotides comprised in the deposited strains, as well as the amino acid sequence of any polypeptide encoded thereby are controlling in the event of any conflict with any description of sequences herein.

A natural mutant of *Streptomyces aizunensis* (NRRL B-11277), referred to as strain [C03]023 producing Compound 2(a) and used to produce the

compounds of Formula I and Formula II was deposited with the International Depositary Authority of Canada, Bureau of Microbiology, Health Canada, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2 on August 7, 2003 and was assigned deposit accession number IDAC 070803-1.

Another mutant of *Streptomyces aizunensis* (NRRL B-11277), referred to as strain [C03U03]023 producing Compound 2(a) and used to produce the compounds of Formula I and Formula II was deposited with the International Depositary Authority of Canada, Bureau of Microbiology, Health Canada, 1015 Arlington Street, Winnipeg, Manitoba, Canada R3E 3R2 on December 23, 2003 and was assigned deposit accession number IDAC 231203-02.

The deposited cosmids and strains [C03]023 and [C03U03]023 (the deposited stains) have been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The deposited strains will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposited strains are provided merely as convenience to those skilled in the art and are not an admission that a deposit is required for enablement. A license may be required to make, use or sell the deposited strains, and compounds derived there from, and no such license is hereby granted.

The order and relative position of the 38 open reading frames representing the proteins of the biosynthetic locus for compound 2(a) (compound 2(a) ORFs) are illustrated schematically in Figure 1. The top line in Figure 1 provides a scale in base pairs. The gray bars depict the two DNA contigs that cover the compound 2(a) locus. The empty arrows represent the 38 open reading frames of the compound 2(a) biosynthetic locus. The black arrows represent the five deposited cosmid clones covering the entire compound 2(a) locus.

One aspect of the present invention is an isolated, purified, or enriched nucleic acid comprising one of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, the sequences complementary thereto, or a fragment comprising at least 100, 200, 300, 400, 500, 600, 700, 800 or more consecutive bases of one of the sequences of

SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 or the sequences complementary thereto. The isolated, purified or enriched nucleic acids may comprise DNA, including cDNA, genomic DNA, and synthetic DNA. The DNA may be double stranded or single stranded, and if single stranded may be the coding (sense) or non-coding (anti-sense) strand. Alternatively, the isolated, purified or enriched nucleic acids may comprise RNA.

As discussed in more detail below, the isolated, purified or enriched nucleic acids of one of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 may be used to prepare one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, respectively, or fragments comprising at least 50, 75, 100, 200, 300, 500 or more consecutive amino acids of one of the polypeptides of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 5 7, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77.

Accordingly, another aspect of the present invention is an isolated, purified or enriched nucleic acid which encodes one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200, 300 or more consecutive amino acids of one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77. The coding sequences of these nucleic acids may be identical to one of the coding sequences of one of the nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 or a fragment thereof, or may be different coding sequences which encode one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50,

75, 100, 150, 200, 300 consecutive amino acids of one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 as a result of the redundancy or degeneracy of the genetic code. The genetic code is well known to those of skill in the art and can be obtained, for example, from Stryer, *Biochemistry*, 3rd edition, W. H. Freeman & Co., New York.

The isolated, purified or enriched nucleic acid which encodes one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 may include, but is not limited to: (1) only the coding sequences of one of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78; (2) the coding sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and additional coding sequences, such as leader sequences or proprotein; and (3) the coding sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 and non-coding sequences, such as non-coding sequences 5' and/or 3' of the coding sequence. Thus, as used herein, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide that includes only coding sequence for the polypeptide as well as a polynucleotide that includes additional coding and/or non-coding sequence.

The invention relates to polynucleotides based on SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78 but having polynucleotide changes that are "silent", for example changes which do not alter the amino acid sequence encoded by the polynucleotides of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78. The invention also relates to polynucleotides which have nucleotide changes which result in amino acid substitutions, additions, deletions, fusions and

truncations of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77. Such nucleotide changes may be introduced using techniques such as site directed mutagenesis, random chemical mutagenesis; exonuclease III deletion, and other recombinant DNA techniques.

The isolated, purified or enriched nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, the sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive bases of one of the sequence of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, or the sequences complementary thereto may be used as probes to identify and isolate DNAs encoding the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 respectively. In such procedures, a genomic DNA library is constructed from a sample microorganism or a sample containing a microorganism capable of producing a polyketide. The genomic DNA library is then contacted with a probe comprising a coding sequence or a fragment of the coding sequence, encoding one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, or a fragment thereof under conditions which permit the probe to specifically hybridize to sequences complementary thereto. In a preferred embodiment, the probe is an oligonucleotide of about 10 to about 30 nucleotides in length designed based on a nucleic acid of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76 or 78. Genomic DNA clones which hybridize to the probe are then detected and isolated. Procedures for preparing and identifying DNA clones of interest are disclosed in Ausubel *et al.*, Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997; and Sambrook *et al.*, Molecular Cloning: A Laboratory

Manual 2d Ed., Cold Spring Harbor Laboratory Press, 1989. In another embodiment, the probe is a restriction fragment or a PCR amplified nucleic acid derived from SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78.

The isolated, purified or enriched nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, the sequences complementary thereto, or a fragment comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400 or 500 consecutive bases of one of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, or the sequences complementary thereto may be used as probes to identify and isolate related nucleic acids. In some embodiments, the related nucleic acids may be genomic DNAs (or cDNAs) from potential polyketide producers. In such procedures, a nucleic acid sample containing nucleic acids from a potential polyketide producer is contacted with the probe under conditions that permit the probe to specifically hybridize to related sequences. The nucleic acid sample may be a genomic DNA (or cDNA) library from the potential polyketide-producer. Hybridization of the probe to nucleic acids is then detected using any of the methods described above.

Hybridization may be carried out under conditions of low stringency, moderate stringency or high stringency. As an example of nucleic acid hybridization, a polymer membrane containing immobilized denatured nucleic acids is first prehybridized for 30 minutes at 45 °C in a solution consisting of 0.9 M NaCl, 50 mM NaH_2PO_4, pH 7.0, 5.0 mM Na_2EDTA, 0.5% SDS, 10X Denhardt's, and 0.5 mg/ml polyriboadenylic acid. Approximately 2 x 10^7 cpm (specific activity 4-9 x 10^8 cpm/ug) of ^{32}P end-labeled oligonucleotide probe are then added to the solution. After 12-16 hours of incubation, the membrane is washed for 30 minutes at room temperature in 1X SET (150 mM NaCl, 20 mM Tris hydrochloride, pH 7.8, 1 mM Na_2EDTA) containing 0.5% SDS, followed by a 30 minute wash in fresh 1X SET at Tm-10°C for the

oligonucleotide probe where Tm is the melting temperature. The membrane is then exposed to autoradiographic film for detection of hybridization signals.

By varying the stringency of the hybridization conditions used to identify nucleic acids, such as genomic DNAs or cDNAs, which hybridize to the detectable probe, nucleic acids having different levels of homology to the probe can be identified and isolated. Stringency may be varied by conducting the hybridization at varying temperatures below the melting temperatures of the probes. The melting temperature of the probe may be calculated using the following formulas:

For oligonucleotide probes between 14 and 70 nucleotides in length the melting temperature (Tm) in degrees Celcius may be calculated using the formula: Tm=81.5+16.6(log [Na+]) + 0.41(fraction G+C)-(600/N) where N is the length of the oligonucleotide.

If the hybridization is carried out in a solution containing formamide, the melting temperature may be calculated using the equation Tm=81.5+16.6(log [Na +]) + 0.41(fraction G + C)-(0.63% formamide)-(600/N) where N is the length of the probe.

Prehybridization may be carried out in 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 0.1 mg/ml denatured fragmented salmon sperm DNA or 6X SSC, 5X Denhardt's reagent, 0.5% SDS, 0.1 mg/ml denatured fragmented salmon sperm DNA, 50% formamide. The composition of the SSC and Denhardt's solutions are listed in Sambrook et al., *supra*.

Hybridization is conducted by adding the detectable probe to the hybridization solutions listed above. Where the probe comprises double stranded DNA, it is denatured by incubating at elevated temperatures and quickly cooling before addition to the hybridization solution. It may also be desirable to similarly denature single stranded probes to eliminate or diminish formation of secondary structures or oligomerization. The filter is contacted with the hybridization solution for a sufficient period of time to allow the probe to hybridize to cDNAs or genomic DNAs containing sequences complementary thereto or homologous thereto. For probes over 200 nucleotides in length, the hybridization may be carried out at 15-25 °C below the Tm. For shorter probes, such as oligonucleotide probes, the hybridization

may be conducted at 5-10 °C below the Tm. Preferably, the hybridization is conducted in 6X SSC, for shorter probes. Preferably, the hybridization is conducted in 50% formamide containing solutions, for longer probes. All the foregoing hybridizations would be considered to be examples of hybridization performed under conditions of high stringency.

Following hybridization, the filter is washed for at least 15 minutes in 2X SSC, 0.1% SDS at room temperature or higher, depending on the desired stringency. The filter is then washed with 0.1X SSC, 0.5% SDS at room temperature (again) for 30 minutes to 1 hour. Nucleic acids which have hybridized to the probe are identified by conventional autoradiography and non-radioactive detection methods.

The above procedure may be modified to identify nucleic acids having decreasing levels of homology to the probe sequence. For example, to obtain nucleic acids of decreasing homology to the detectable probe, less stringent conditions may be used. For example, the hybridization temperature may be decreased in increments of 5 °C from 68 °C to 42 °C in a hybridization buffer having a Na+ concentration of approximately 1M. Following hybridization, the filter may be washed with 2X SSC, 0.5% SDS at the temperature of hybridization. These conditions are considered to be "moderate stringency" conditions above 50°C and "low stringency" conditions below 50°C. A specific example of "moderate stringency" hybridization conditions is when the above hybridization is conducted at 55°C. A specific example of "low stringency" hybridization conditions is when the above hybridization is conducted at 45°C.

Alternatively, the hybridization may be carried out in buffers, such as 6X SSC, containing formamide at a temperature of 42 °C. In this case, the concentration of formamide in the hybridization buffer may be reduced in 5% increments from 50% to 0% to identify clones having decreasing levels of homology to the probe. Following hybridization, the filter may be washed with 6X SSC, 0.5% SDS at 50 °C. These conditions are considered to be "moderate stringency" conditions above 25% formamide and "low stringency" conditions below 25% formamide. A specific example of "moderate stringency" hybridization conditions is when the above hybridization is conducted at 30% formamide. A specific example of "low stringency"

hybridization conditions is when the above hybridization is conducted at 10% formamide. Nucleic acids which have hybridized to the probe are identified by conventional autoradiography and non-radioactive detection methods.

The preceding methods may be used to isolate nucleic acids having at least 97%, at least 95%, at least 90%, at least 85%, at least 80%, or at least 70% sequence identity to a nucleic acid sequence selected from the group consisting of the sequences of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, fragments comprising at least 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, 150, 200, 300, 400, or 500 consecutive bases thereof, and the sequences complementary thereto. The isolated nucleic acid may have a coding sequence that is a naturally occurring allelic variant of one of the coding sequences described herein. Such allelic variant may have a substitution, deletion or addition of one or more nucleotides when compared to the nucleic acids of SEQ ID NOS: 3, 5, 7, 9, 11, 13, 15, 17, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, or the sequences complementary thereto.

Additionally, the above procedures may be used to isolate nucleic acids which encode polypeptides having at least 99%, at least 95%, at least 90%, at least 85%, at least 80%, or at least 70% identity to a polypeptide having the sequence of one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200, 300 consecutive amino acids thereof as determined using the BLASTP version 2.2.2 algorithm with default parameters.

Another aspect of the present invention is an isolated or purified polypeptide comprising the sequence of one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof. As discussed herein, such polypeptides may be obtained by inserting a nucleic acid encoding the polypeptide into a vector such that the coding sequence is operably linked to a sequence capable of driving the expression of the

encoded polypeptide in a suitable host cell. For example, the expression vector may comprise a promoter, a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for modulating expression levels, an origin of replication and a selectable marker.

Promoters suitable for expressing the polypeptide or fragment thereof in bacteria include the *E.coli lac* or *trp* promoters, the lacI promoter, the lacZ promoter, the T3 promoter, the T7 promoter, the gpt promoter, the lambda P_R promoter, the lambda P_L promoter, promoters from operons encoding glycolytic enzymes such as 3-phosphoglycerate kinase (PGK), and the acid phosphatase promoter. Fungal promoters include the α factor promoter. Eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, heat shock promoters, the early and late SV40 promoter, LTRs from retroviruses, and the mouse metallothionein-I promoter. Other promoters known to control expression of genes in prokaryotic or eukaryotic cells or their viruses may also be used.

Mammalian expression vectors may also comprise an origin of replication, any necessary ribosome binding sites, a polyadenylation site, splice donors and acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences. In some embodiments, DNA sequences derived from the SV40 splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Vectors for expressing the polypeptide or fragment thereof in eukaryotic cells may also contain enhancers to increase expression levels. Enhancers are cis-acting elements of DNA, usually from about 10 to about 300 bp in length that act on a promoter to increase its transcription. Examples include the SV40 enhancer on the late side of the replication origin bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and the adenovirus enhancers.

In addition, the expression vectors preferably contain one or more selectable marker genes to permit selection of host cells containing the vector. Examples of selectable markers that may be used include genes encoding dihydrofolate reductase or genes conferring neomycin resistance for

eukaryotic cell culture, genes conferring tetracycline or ampicillin resistance in *E. coli*, and the *S. cerevisiae* TRP1 gene.

In some embodiments, the nucleic acid encoding one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof is assembled in appropriate phase with a leader sequence capable of directing secretion of the translated polypeptides or fragments thereof. Optionally, the nucleic acid can encode a fusion polypeptide in which one of the polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof is fused to heterologous peptides or polypeptides, such as N-terminal identification peptides which impart desired characteristics such as increased stability or simplified purification or detection.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is ligated to the desired position in the vector following digestion of the insert and the vector with appropriate restriction endonucleases. Alternatively, appropriate restriction enzyme sites can be engineered into a DNA sequence by PCR. A variety of cloning techniques are disclosed in Ausbel *et al.* Current Protocols in Molecular Biology, John Wiley 503 Sons, Inc. 1997 and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual 2d Ed., Cold Spring Harbour Laboratory Press, 1989. Such procedures and others are deemed to be within the scope of those skilled in the art.

The vector may be, for example, in the form of a plasmid, a viral particle, or a phage. Other vectors include derivatives of chromosomal, nonchromosomal and synthetic DNA sequences, viruses, bacterial plasmids, phage DNA, baculovirus, yeast plasmids, vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, and pseudorabies. A variety of cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook *et al.*,

Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989).

Particular bacterial vectors which may be used include the commercially available plasmids comprising genetic elements of the well known cloning vector pBR322 (ATCC 37017), pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden), pGEM1 (Promega Biotec, Madison, WI, USA) pQE70, pQE60, pQE-9 (Qiagen), pD10, phiX174, pBluescript™ II KS, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene), ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia), pKK232-8 and pCM7. Particular eukaryotic vectors include pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, and pSVL (Pharmacia). However, any other vector may be used as long as it is replicable and stable in the host cell.

The host cell may be any of the host cells familiar to those skilled in the art, including prokaryotic cells or eukaryotic cells. As representative examples of appropriate hosts, there may be mentioned: bacteria cells, such as *E. coli*, *Streptomyces lividans*, *Streptomyces griseofuscus*, *Streptomyces ambofaciens*, *Bacillus subtilis*, *Salmonella typhimurium* and various species within the genera *Pseudomonas*, *Streptomyces*, *Bacillus*, and *Staphylococcus*, fungal cells, such as yeast, insect cells such as *Drosophila S2* and *Spodoptera Sf9*, animal cells such as CHO, COS or Bowes melanoma, and adenoviruses. The selection of an appropriate host is within the abilities of those skilled in the art.

The vector may be introduced into the host cells using any of a variety of techniques, including electroporation transformation, transfection, transduction, viral infection, gene guns, or Ti-mediated gene transfer. Where appropriate, the engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes of the present invention. Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter may be induced by appropriate means (e.g., temperature shift or chemical induction) and the cells may be cultured for an additional period to allow them to produce the desired polypeptide or fragment thereof.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract is retained for further purification. Microbial cells employed for expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known to those skilled in the art. The expressed polypeptide or fragment thereof can be recovered and purified from recombinant cell cultures by methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the polypeptide. If desired, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Various mammalian cell culture systems can also be employed to express recombinant protein. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblasts (described by Gluzman, Cell, 23:175(1981)), and other cell lines capable of expressing proteins from a compatible vector, such as the C127, 3T3, CHO, HeLa and BHK cell lines. The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Polypeptides of the invention may or may not also include an initial methionine amino acid residue.

Alternatively, the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof can be synthetically produced by conventional peptide synthesizers. In other embodiments, fragments or portions of the polynucleotides may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides.

Cell-free translation systems can also be employed to produce one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof using mRNAs transcribed from a DNA construct comprising a promoter operably linked to a nucleic acid encoding the polypeptide or fragment thereof. In some embodiments, the DNA construct may be linearized prior to conducting an *in vitro* transcription reaction. The transcribed mRNA is then incubated with an appropriate cell-free translation extract, such as a rabbit reticulocyte extract, to produce the desired polypeptide or fragment thereof.

The present invention also relates to variants of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof. The term "variant" includes derivatives or analogs of these polypeptides. In particular, the variants may differ in amino acid sequence from the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

The variants may be naturally occurring or created *in vitro*. In particular, such variants may be created using genetic engineering techniques such as site directed mutagenesis, random chemical mutagenesis, exonuclease III deletion procedures, and standard cloning techniques. Alternatively, such variants, fragments, analogs, or derivatives may be created using chemical synthesis or modification procedures.

Other methods of making variants are also familiar to those skilled in the art. These include procedures in which nucleic acid sequences obtained from natural isolates are modified to generate nucleic acids that encode polypeptides having characteristics which enhance their value in industrial or laboratory applications. In such procedures, a large number of variant sequences having one or more nucleotide differences with respect to the

sequence obtained from the natural isolate are generated and characterized. Preferably, these nucleotide differences result in amino acid changes with respect to the polypeptides encoded by the nucleic acids from the natural isolates.

For example, variants may be created using error prone PCR. In error prone PCR, DNA amplification is performed under conditions where the fidelity of the DNA polymerase is low, such that a high rate of point mutation is obtained along the entire length of the PCR product. Error prone PCR is described in Leung, D.W., *et al.*, Technique, 1:11-15 (1989) and Caldwell, R. C. & Joyce G.F., PCR Methods Applic., 2:28-33 (1992). Variants may also be created using site directed mutagenesis to generate site-specific mutations in any cloned DNA segment of interest. Oligonucleotide mutagenesis is described in Reidhaar-Olson, J.F. & Sauer, R.T., *et al.*, Science, 241:53-57 (1988). Variants may also be created using directed evolution strategies such as those described in US patent nos. 6,361,974 and 6,372,497. The variants of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75 and 77 may be variants in which one or more of the amino acid residues of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75 or 77 are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code.

Conservative substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the following replacements: replacements of an aliphatic amino acid such as Ala, Val, Leu and Ile with another aliphatic amino acid; replacement of a Ser with a Thr or vice versa; replacement of an acidic residue such as Asp or Glu with another acidic residue; replacement of a residue bearing an amide group, such as Asn or Gln, with another residue bearing an amide group; exchange of a basic residue such as Lys or Arg with

another basic residue; and replacement of an aromatic residue such as Phe or Tyr with another aromatic residue.

Other variants are those in which one or more of the amino acid residues of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 include a substituent group. Still other variants are those in which the polypeptide is associated with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol). Additional variants are those in which additional amino acids are fused to the polypeptide, such as leader sequence, a secretory sequence, a proprotein sequence or a sequence that facilitates purification, enrichment, or stabilization of the polypeptide.

In some embodiments, the fragments, derivatives and analogs retain the same biological function or activity as the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77. In other embodiments, the fragment, derivative or analogue includes a fused heterologous sequence that facilitates purification, enrichment, detection, stabilization or secretion of the polypeptide that can be enzymatically cleaved, in whole or in part, away from the fragment, derivative or analogue.

Another aspect of the present invention are polypeptides or fragments thereof which have at least 70%, at least 80%, at least 85%, at least 90%, or more than 95% identity to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75 and 77 or a fragment comprising at least 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof. It will be appreciated that amino acid "identity" includes conservative substitutions such as those described above.

The polypeptides or fragments having homology to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or a fragment comprising at least 50, 75, 100, 150, 200 or 300

consecutive amino acids thereof may be obtained by isolating the nucleic acids encoding them using the techniques described above.

Alternatively, the homologous polypeptides or fragments may be obtained through biochemical enrichment or purification procedures. The sequence of potentially homologous polypeptides or fragments may be determined by proteolytic digestion, gel electrophoresis and/or microsequencing. The sequence of the prospective homologous polypeptide or fragment can be compared to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or a fragment comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof.

The polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments, derivatives or analogs thereof comprising at least 40, 50, 75, 100, 150, 200 or 300 consecutive amino acids thereof invention may be used in a variety of applications. For example, the polypeptides or fragments, derivatives or analogs thereof may be used to catalyze biochemical reactions as described elsewhere in the specification.

The polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments, derivatives or analogues thereof comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof, may also be used to generate antibodies which bind specifically to the polypeptides or fragments, derivatives or analogues. The antibodies generated from SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 may be used to determine whether a biological sample contains *Streptomyces aizunensis* or a related microorganism.

In such procedures, a biological sample is contacted with an antibody capable of specifically binding to one of the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45,

47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. The ability of the biological sample to bind to the antibody is then determined. For example, binding may be determined by labeling the antibody with a detectable label such as a fluorescent agent, an enzymatic label, or a radioisotope. Alternatively, binding of the antibody to the sample may be detected using a secondary antibody having such a detectable label thereon. A variety of assay protocols which may be used to detect the presence of a polyketide-producer or of *Streptomyces aizunensis* or of polypeptides related to SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 in a sample are familiar to those skilled in the art. Particular assays include ELISA assays, sandwich assays, radioimmunoassays, and Western Blots. Alternatively, antibodies generated from SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 may be used to determine whether a biological sample contains related polypeptides that may be involved in the biosynthesis of polyketides.

Polyclonal antibodies generated against the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptide itself. In this manner, even a sequence encoding only a fragment of the polypeptide can be used to generate antibodies that may bind to the whole native polypeptide. Such antibodies can then be used to isolate the polypeptide from cells expressing that polypeptide.

For preparation of monoclonal antibodies, any technique that provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kholer and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et

al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof. Alternatively, transgenic mice may be used to express humanized antibodies to these polypeptides or fragments thereof.

Antibodies generated against the polypeptides of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77 or fragments comprising at least 5, 10, 15, 20, 25, 30, 35, 40, 50, 75, 100, or 150 consecutive amino acids thereof may be used in screening for similar polypeptides from a sample containing organisms or cell-free extracts thereof. In such techniques, polypeptides from the sample are contacted with the antibodies and those polypeptides which specifically bind the antibody are detected. Any of the procedures described above may be used to detect antibody binding. One such screening assay is described in "Methods for measuring Cellulase Activities", Methods in Enzymology, Vol 160, pp. 87-116.

In order to identify the function of the genes in the compound 2(a) locus, ORFs 1 to 38 were compared, using the BLASTP version 2.2.1 algorithm with the default parameters, to sequences in the National Center for Biotechnology Information (NCBI) nonredundant protein database and the DECIPHER® database of microbial genes, pathways and natural products (Ecopia BioSciences Inc. St.-Laurent, QC, Canada).

The accession numbers of the top GenBank hits of this Blast analysis are presented in Table 1 along with the corresponding E values. The E value relates the expected number of chance alignments with an alignment score at least equal to the observed alignment score. An E value of 0.00 indicates a perfect homolog. The E values are calculated as described in Altschul *et al. J.*

Mol. Biol., 215, 403-410 (1990). The E value assists in the determination of whether two sequences display sufficient similarity to justify an inference of homology.

Table 1

ORF	Family	#aa	GenBank homology	probability	% identity	% similarity	proposed function of GenBank match
1		719	T35189,719aa	1E-200	556/705 (78.7%)	582/705 (82.55%)	helicase, Streptomyces coelicolor
			BAC17778.1,686aa	1E-165	340/700 (48.57%)	407/700 (58.14%)	helicase, Corynebacterium efficiens
			NP_600121.1,683aa	1E-161	334/701 (47.65%)	412/701 (58.77%)	helicase, Corynebacterium glutamicum
2	TESA	253	BAB69315.1,255aa	2E-82	142/243 (58.44%)	185/243 (76.13%)	thioesterase,Streptomyces avermitilis
			CAC20922.1,255aa	3E-78	145/247 (58.7%)	180/247 (72.87%)	PimI thioesterase, Streptomyces natalensis
			AAF71777.1,251aa	2E-73	135/244 (55.33%)	173/244 (70.9%)	NysE thioesterase, Streptomyces noursei
3	REGD	956	AAC68887.1,928aa	1E-132	336/959 (35.04%)	472/959 (49.22%)	transcriptional activator,Streptomyces venezuelae
			AAM88362.1,945aa	1E-131	331/957 (34.59%)	468/957 (48.9%)	NbmM regulator,Streptomyces narbonensis
			BAA84600.1,949aa	1E-127	339/965 (35.13%)	451/965 (46.74%)	response regulator,Streptomyces avermitilis
4	RREB	201	NP_629592.1,224aa	8E-49	106/204 (51.96%)	140/204 (68.63%)	response regulator,Streptomyces coelicolor
			CAA74720.1,217aa	8E-47	100/202 (49.5%)	138/202 (68.32%)	response regulator,Streptomyces reticuli
			NP_642485.1,213aa	1E-43	96/201 (47.76%)	132/201 (65.67%)	regulator,Xanthomonas axonopodis
5	SPKK	416	NP_628447.1,428aa	7E-39	116/312 (37.18%)	163/312 (52.24%)	kinase,Streptomyces coelicolor
			CAA74719.1,398aa	2E-37	113/304 (37.17%)	157/304 (51.64%)	kinase,Streptomyces reticuli
			CAC32293.1,404aa	6E-37	109/267 (40.82%)	139/267 (52.06%)	kinase,Streptomyces coelicolor
6	UNEW	186	NP_628531.1,185aa	0.002	30/102 (29.41%)	50/102 (49.02%)	membrane protein,Streptomyces coelicolor
7	UNFI	163	CAB10923.1,177aa	2E-27	67/162 (41.36%)	97/162 (59.88%)	hypothetical protein,Mycobacterium tuberculosis
			ZP_00059442.1,172aa	4E-24	66/177 (37.29%)	98/177 (55.37%)	hypothetical protein,Thermobifida fusca
			NP_644099.1,158aa	1E-08	35/107 (32.71%)	58/107 (54.21%)	hypothetical protein,Xanthomonas axonopodis
8	UNEX	514	E70508,487aa	4E-41	145/494 (29.35%)	210/494 (42.51%)	hypothetical protein,Mycobacterium tuberculosis
			ZP_00059443.1,554aa	5E-39	155/516 (30.04%)	216/516 (41.86%)	hypothetical protein,Thermobifida fusca
			NP_280206.1,514aa	1E-07	107/475 (22.53%)	169/475 (35.58%)	hypothetical protein,Halobacterium sp.
9	GTFA	367	AAM94798.1,376aa	4E-69	155/370 (41.89%)	205/370 (55.41%)	CalG3 glycosyltransferase,Micromonospora echinospora
			CAC16413.2,382aa	2E-60	150/373 (40.21%)	187/373 (50.13%)	glycosyltransferase,Streptomyces olivaceus
			AAF01811.1,390aa	5E-54	138/374 (36.9%)	179/374 (47.86%)	glycosyl transferase,Streptomyces nogalater

10	PKSH	8147	AAG23283.1.4928aa	1E-200	2487/5027 (49.47%)	3043/5027 (60.53%)	polyketide synthase,Saccharopolyspora spinosa
		BAB69303.1.6048aa	1E-200	2452/4407 (55.64%)	2853/4407 (64.74%)	polyketide synthase,Streptomyces avermitilis
		AAF71766.1.9477aa	1E-200	2489/4970 (50.08%)	3044/4970 (61.25%)	Nysl polyketide synthase,Streptomyces noursei
11	PKSH	3428	AAF71766.1.9477aa	1E-200	1763/3432 (51.37%)	2086/3432 (60.78%)	Nysl polyketide synthase,Streptomyces noursei
		AAK73501.1.9510aa	1E-200	1738/3394 (51.21%)	2065/3394 (60.84%)	Amphl polyketide synthase,Streptomyces nodosus
		CAC20921.1.9507aa	1E-200	1729/3385 (51.08%)	2050/3385 (60.56%)	PlmS2 polyketide synthase,Streptomyces natalensis
12	PKSH	6751	AAF71766.1.9477aa	1E-200	2992/5949 (50.29%)	3662/5949 (61.56%)	Nysl polyketide synthase,Streptomyces noursei
		AAK73501.1.9510aa	1E-200	2961/5904 (50.15%)	3650/5904 (61.82%)	Amphl polyketide synthase,Streptomyces nodosus
		CAC20921.1.9507aa	1E-200	2962/5917 (50.06%)	3652/5917 (61.72%)	PlmS2 polyketide synthase,Streptomyces natalensis
13	PKSH	1657	AAF71775.1.3192aa	1E-200	781/1553 (50.29%)	939/1553 (60.4%)	NysB polyketide synthase,Streptomyces noursei
		BAB69196.1.3613aa	1E-200	775/1562 (49.62%)	936/1562 (59.92%)	polyketide synthase,Streptomyces avermitilis
		AAG23266.1.3170aa	1E-200	775/1572 (49.3%)	941/1572 (59.86%)	polyketide synthase,Saccharopolyspora spinosa
14	PKSH	5207	BAB69303.1.6048aa	1E-200	2713/5239 (51.79%)	3215/5239 (61.37%)	polyketide synthase,Streptomyces avermitilis
		CAC20931.1.6797aa	1E-200	2651/5183 (51.15%)	3187/5183 (61.49%)	PlmS1polyketide synthase,Streptomyces natalensis
		AAK73514.1.10917aa	1E-200	2047/4174 (49.04%)	2494/4174 (59.75%)	AmphC polyketide synthase,Streptomyces nodosus
15	PKSH	5432	AAK73514.1.10917aa	1E-200	2814/5447 (51.66%)	3377/5447 (62%)	AmphC polyketide synthase,Streptomyces nodosus
		AAF71776.1.11098aa	1E-200	2836/5548 (51.12%)	3375/5548 (60.83%)	NysC polyketide synthase,Streptomyces noursei
		CAC20931.1.6797aa	1E-200	2824/5426 (52.06%)	3378/5426 (62.26%)	PlmS1polyketide synthase,Streptomyces natalensis
16	PKSH	3227	AAF71775.1.3192aa	1E-200	1628/3207 (50.76%)	1957/3207 (61.02%)	NysB polyketide synthase,Streptomyces noursei

		AAF82408.1|4150aa	1E-200	1643/3237 (50.76%)	1957/3237 (60.46%)	deoxyoleandolide synthase,Streptomyces antibioticus
		BAB69307.1|3352aa	1E-200	1612/3170 (50.85%)	1948/3170 (61.45%)	polyketide synthase,Streptomyces avermitilis
17	PKSH	7510	AAK73502.1|5644aa	1E-200	2761/5719 (48.28%)	3366/5719 (58.86%)	AmphJ polyketide synthase,Streptomyces nodosus
		AAF71776.1|1109aa	1E-200	2313/4464 (51.81%)	2755/4464 (61.72%)	NysC polyketide synthase,Streptomyces noursei
		CAA60460.1|8563aa	1E-200	2448/5643 (43.38%)	3074/5643 (54.47%)	polyketide synthase,Streptomyces hygroscopicus
18	PKSH	3872	AAK73514.1|10917aa	1E-200	1913/3588 (53.32%)	2273/3588 (63.35%)	AmphC polyketide synthase,Streptomyces nodosus
		AAF71776.1|1109aa	1E-200	1907/3684 (51.76%)	2280/3684 (61.89%)	NysC polyketide synthase,Streptomyces noursei
		CAC20931.1|6797aa	1E-200	1879/3564 (52.72%)	2241/3564 (62.88%)	PimS1 polyketide synthase,Streptomyces natalensis
19	AYTF	338	D83961.1|313aa	1E-09	72/294 (24.49%)	118/294 (40.14%)	malonyl CoA-ACP transacylase,Bacillus halodurans
		AAL20123.1|309aa	1E-08	73/303 (24.09%)	120/303 (39.6%)	malonyl-CoA-ACP transacylase,Salmonella typhimurium
		AAK60008.1|316aa	1E-07	74/286 (25.87%)	110/286 (38.46%)	malonyl-CoA-ACP transacylase,Streptomyces aureofaciens
20	MEAY	283	AD2393.1|275aa	6E-11	60/220 (27.27%)	97/220 (44.09%)	hypothetical protein,Nostoc sp.
		S76277.1|284aa	2E-08	70/255 (27.45%)	112/255 (43.92%)	hypothetical protein,Synechocystis sp.
		ZP_00019722.1|251aa	6E-08	58/224 (25%)	99/224 (44.2%)	hypothetical protein,Chloroflexus aurantiacus
21	ABCD	329	ZP_00080468.1|308aa	1E-54	142/334 (42.51%)	176/334 (52.69%)	hypothetical protein,Geobacter metallireducens
		D72257.1|327aa	5E-52	131/330 (39.7%)	186/330 (56.36%)	hypothetical protein,Thermotoga maritima
		NP_573812.1|321aa	5E-49	121/327 (37%)	176/327 (53.82%)	daunorubicin resistance protein,Pyrococcus furiosus
22	DEPL	317	CAA07388.1|305aa	4E-73	152/290 (52.41%)	173/290 (59.66%)	StrL,Streptomyces glaucescens
		AAF59696.1|294aa	5E-65	139/285 (48.77%)	165/285 (57.89%)	4-ketoreductase,Streptomyces antibioticus
		AAF01815.1|291aa	8E-63	136/289 (47.06%)	161/289 (55.71%)	dTDP-4-dehydrorhamnose reductase,Streptomyces nogalater
23	EPIM	204	CAA44442.1|200aa	1E-56	108/195 (55.38%)	137/195 (70.26%)	epimerase,Streptomyces griseus

		CAA55578.1,200aa	2E-55	108/191 (56.54%)	129/191 (67.54%)		epimerase,Streptomyces glaucescens
		AAG29805.1,198aa	1E-51	104/188 (55.32%)	121/188 (64.36%)		epimerase,Streptomyces rishiriensis
24	NUTA	328	CAA65514.1,355aa	1E-125	215/328 (65.55%)	263/328 (80.18%)	sugaractivating enzyme,Streptomyces griseus
			BAC55207.1,350aa	1E-122	217/328 (66.16%)	261/328 (79.57%)	glucose-1-phosphate thymidylyltransferase,Streptomyces sp.
			AAF59834.1,356aa	1E-119	214/329 (65.06%)	260/329 (79.03%)	dTDP-D-glucose synthase,Streptomyces antibioticus
25	DEPA	328	NP_625052.1,324aa	1E-108	201/318 (63.21%)	218/318 (68.55%)	putative dehydratase,Streptomyces coelicolor
			CAA07755.1,331aa	1E-107	200/317 (63.09%)	218/317 (68.77%)	dTDP-glucose 4,6-dehydratase,Streptomyces argillaceus
			AAF82606.1,317aa	1E-105	191/318 (60.06%)	214/318 (67.3%)	dTDP-glucose 4,6-dehydratase,Streptomyces rimosus
26	TESA	214	BAB69315.1,255aa	7E-22	74/239 (30.96%)	97/239 (40.59%)	thioesterase,Streptomyces avermitilis
			T17413,281aa	3E-17	73/242 (30.17%)	95/242 (39.26%)	thioesterase,Streptomyces venezuelae
			AAC01736.1,254aa	3E-15	61/225 (27.11%)	88/225 (39.11%)	thioesterase,Amycolatopsis mediterranei
27	CALB	470	ZP_00025699.1,510aa	1E-27	116/466 (24.89%)	188/466 (40.34%)	hypothetical protein,Ralstonia metallidurans
			ZP_00006768.1,501aa	5E-22	125/474 (26.37%)	192/474 (40.51%)	hypothetical protein,Rhodobacter sphaeroides
			G87227,548aa	1e-20	120/495 (24.24%)	195/495 (39.39%)	acyl-CoA synthase,Mycobacterium leprae
28	TMOA	553	CAB76876.1,565aa	1E-200	318/522 (60.92%)	383/522 (73.37%)	amino oxidase,Streptomyces coelicolor
			ZP_00086824.1,560aa	1E-172	280/521 (53.74%)	369/521 (70.83%)	hypothetical protein,Pseudomonas fluorescens
			ZP_00126831.1,559aa	1E-171	280/521 (53.74%)	370/521 (71.02%)	tryptophan monooxygenase,Pseudomonas syringae
29	PPTF	231	CAA19952.1,226aa	3E-50	115/226 (50.88%)	132/226 (58.41%)	hypothetical protein,Streptomyces coelicolor
			AAG43513.1,246aa	1E-43	105/228 (46.05%)	127/228 (55.7%)	phosphopantetheinyl transferase,Streptomyces verticillus
			BAA22407.1,208aa	5E-36	91/214 (42.52%)	109/214 (50.93%)	hypothetical protein,Streptomyces sp.
30	UNAK	306	CAA19951.1,295aa	2E-97	169/275 (61.45%)	195/275 (70.91%)	hypothetical protein,Streptomyces coelicolor
			AAL15596.1,293aa	1E-91	163/269 (60.59%)	190/269 (70.63%)	Slm18,Streptomyces antibioticus
			NP_217311.1,324aa	8E-89	159/276 (57.61%)	187/276 (67.75%)	hypothetical protein,Mycobacterium tuberculosis
31	REGD	998	AAC68887.1,928aa	1E-113	331/1014 (32.64%)	445/1014 (43.89%)	transcriptional activator,Streptomyces venezuelae

			AAM88362.1,945aa	1E-110	329/1007 (32.08%)	438/1007 (43.5%)	NbrnM regulator,Streptomyces narbonensis
							regulatory protein,Streptomyces
			AAC38065.1,948aa	1E-105	322/1019 (31.6%)	427/1019 (41.9%)	hygroscopicus
32	CITFC	518	NP_629669.1,527aa	1E-200	461/516 (89.34%)	487/516 (94.38%)	carboxyl transferase,Streptomyces coelicolor
							putative decarboxylase,Streptomyces
			AAK06793.1,528aa	1E-200	423/510 (82.94%)	464/510 (90.98%)	antibioticus
			AAD13544.1,524aa	1E-200	411/511 (80.43%)	450/511 (88.06%)	decarboxylase,Streptomyces cyanogenus
33	ADHY	329	CAD55203.1,322aa	1E-140	240/314 (76.43%)	279/314 (88.94%)	putative agmatinase,Streptomyces coelicolor
			ZP_00057179.1,324aa	1E-128	216/312 (69.23%)	260/312 (83.33%)	hypothetical protein,Thermobifida fusca
			BAB96819.1,353aa	1E-119	206/307 (67.1%)	252/307 (82.08%)	guanidinobutyrase,Arthrobacter sp.
34	ADSN	521	AAG29784.1,529aa	6E-86	189/512 (36.91%)	255/512 (49.8%)	ligase,Streptomyces rishiriensis
							amide synthetase,Streptomyces
			AAN66228.1,527aa	1E-81	183/512 (35.74%)	252/512 (49.22%)	roseochromogenes
							aminocoumarin ligase,Streptomyces
			AAG34183.1,519aa	3E-74	186/515 (36.12%)	248/515 (48.16%)	antibioticus
35	AYTP	410	NP_697353.1,425aa	1E-104	193/385 (50.13%)	252/385 (65.45%)	5-aminolevulinic acid synthase,Brucella suis
			AAL52785.1,425aa	1E-103	192/385 (49.87%)	250/385 (64.94%)	5-aminolevulinic acid synthase,Brucella melitensis
			BAB52860.1,425aa	1E-102	191/385 (49.61%)	249/385 (64.68%)	5-aminolevulinic acid synthase,Mesorhizobium loti
36	CALB	506	CAB89029.1,511aa	1E-134	256/505 (50.69%)	319/505 (63.17%)	long-chain-fatty-acid-CoA ligase,Streptomyces coelicolor
			ZP_00059397.1,557aa	1E-122	241/505 (47.72%)	315/505 (62.38%)	hypothetical protein,Thermobifida fusca
			ZP_00105928.1,513aa	7E-78	195/501 (38.92%)	260/501 (51.9%)	hypothetical protein,Nostoc punctiforme
37		217	NP_635504.1,210aa	6E-11	54/139 (38.85%)	77/139 (55.4%)	regulatory protein,Xanthomonas campestris
							response regulator,Halomonas
			BAB84309.1,221aa	6E-11	66/221 (29.86%)	101/221 (45.7%)	halodenitrificans
			NP_631750.1,226aa	8E-11	71/218 (32.57%)	99/218 (45.41%)	response regulator,Streptomyces coelicolor
38		442	CAA19514.1,534aa	1E-116	206/359 (57.38%)	251/359 (69.92%)	hypothetical protein,Streptomyces coelicolor
			ZP_00058746.1,366aa	7E-19	99/358 (27.37%)	145/358 (40.5%)	hypothetical protein,Thermobifida fusca
							hypothetical protein,Mycobacterium
			AAK47101,352aa	3E-08	85/357 (23.81%)	130/357 (36.42%)	tuberculosis

The gene product of each of ORFs 1-38 in the compound 2(a) locus is assigned a protein family based on sequence similarity to the structure of known proteins as determined in Table 1. A putative function is attributed to each gene product of the compound 2(a) locus biosynthetic locus based on the known function of members of the respective protein families. Each protein family is referred to by a four-letter designation used throughout the description and figures. For example, members of protein family ABCD including the gene product of ORF 21 (SEQ ID NO: 43) are transmembrane transporters; members of protein family ADHY including the gene product ORF 33 (SEQ ID NO: 67) are amidinohydrolases; members of protein family ADSN including the gene product of ORF 34 (SEQ ID NO: 69) are adenylation/condensing enzymes; members of protein families AYTF and AYTP including ORFs 19 and 35 (SEQ ID NOS: 39 and 71) are acyltransferases; members of protein family CALB are acyl CoA ligases including ORF 27 and 36 (SEQ ID NO: 55 and 73); members of protein family CTFC including ORF 32 (SEQ ID NO: 65) are carboxyltransferase/decarboxylases; members of protein families DEPA and DEPL including ORFs 25 and 22 (SEQ ID NOS: 51 and 45) are dehydratase/epimerases; members of protein family EPIM including ORF 23 (SEQ ID NO: 47) are epimerises; members of protein family GTFA including ORF 9 (SEQ ID NO: 19) are glycosyl transferases; members of protein family MEAY including ORF 20 (SEQ ID NO: 41) are membrane proteins; members of protein family NUTA including ORF 24 (SEQ ID NO: 49) are nucleotidyltransferases; members of protein family PKSH including ORFs 10, 11, 12, 13, 14, 15, 16, 17 and 18 (SEQ ID NOS: 21, 23, 25, 27, 29, 31, 33, 35 and 37) are polyketide synthase, type I proteins; members of PPTF protein family including ORF 29 (SEQ ID NO: 59) are phosphopantetheinyl transferases; members of protein family REGD including ORFs 3 and 31 (SEQ ID NOS: 6 and 63) are transcriptional regulators; members of protein family RREB including ORF 4 (SEQ ID NO: 8) are response regulators; members of protein family SPKK including ORF 5 (SEQ ID NO: 10) are sensory protein kinases; members of protein family TESA including ORFs 2 and 26 (SEQ ID NOS: 4 and 53) are thioesterases; and members of protein family TMOA including ORF 28 (SEQ ID NO: 57) are monooxygenases. A

more detailed description of the function of each protein family is provided in Table 2. The correlation between structure and function for each protein family is provided in Table 2.

Table 2

Protein Family	Function
ABCD	ABC transporter; ATP-binding cassette transmembrane transporter; includes proteins with similarity to Mdr proteins of mammalian tumor cells that confer resistance to chemotherapeutic agents.
ADHY	amidinohydrolase; agmatine ureohydrolase; hydrolyzes linear amidines; requires manganese for catalysis and contains a conserved His important for catalytic function
ADSN	Adenylating/condensing synthase; amide synthase; enzymes able to activate substrates as acyl adenylates and subsequently transfer the acyl group to an amino group of the acceptor molecule
AYTF	acyltransferase; acyl CoA-acyl carrier protein transacylase; includes malonyl CoA-ACP transacylases
AYTP	acyltransferase; pyridoxal phosphate-dependent; includes 5-aminolevulinate synthase, a glycyl transferase that condenses glycine and succinyl-CoA.
CALB	acyl CoA ligase; shows similarity to plant coumarate CoA ligases, other aryl CoA ligases, yeast CoA synthetase and aminocoumarin ligases.
CTFC	carboxyltransferase/decarboxylase; carboxyltransferase component of acetyl-CoA carboxylase, generally a 2 subunit component, this family consists of a fusion of the beta and alpha subunits (beta-alpha).
DEPA	dehydratase/epimerase; dTDP-glucose 4,6-dehydratases, catalyze the second step in 6-deoxyhexose biosynthesis.
DEPL	dehydratase/epimerase; similar to StrL dTDP-dihydrostreptose synthase; OleU 4-ketoreductase; SnogC putative dTDP-4-dehydrorhamnose reductase
EPIM	epimerase; NDP-hexose epimerase; TDP-4-ketohexose- 3,5-epimerases, convert TDP-4-keto-6-deoxy-D-glucose to TDP-4-keto-6-deoxy-L-mannose (TDP-4-keto-L-rhamnose).
GTFA	glycosyl transferase.
MEAY	membrane protein; putative transporter, permease
NUTA	nucleotidyltransferase; dNDP-glucose synthase; alpha-D-glucose-1-phosphate thymidylyltransferase; catalyze the first step in 6-deoxyhexose biosynthesis.
PKSH	polyketide synthase, type I.
PPTF	phosphopantetheinyl transferases, required for activation of both PKSs and NRPSs from inactive apo forms to active holo forms.
REGD	transcriptional regulator
RREB	response regulator; similar to response regulators that are known to bind DNA and act as transcriptional activators
SPKK	sensory protein kinase.
TESA	thioesterase.
TMOA	monooxygenase; strong similarity to plasmid-encoded tryptophan-2-monooxygenases.

UNAK	unknown; homolog of S. coelicolor hypothetical protein
UNEW	unknown; similar to putative integral membrane protein in S. coelicolor
UNEX	unknown; domain homology to many bacterial putative membrane proteins; contain so-called "bacterial membrane flanked domains" found in an uncharacterised family of membrane proteins that have one to three copies of the domain flanked by transmembrane helices.
UNFI	unknown; similar to putative membrane proteins

Biosynthesis of Compound 2(a) involves the multimodular type I polyketide synthase system (PKS) of ORFs 10 to 18 (SEQ ID NOS: 21, 23, 25, 27, 29, 31, 33, 35 and 37) illustrated in Figure 1. Type I PKSs are large modular proteins that condense acyl thioester units in a sequential manner. PKS systems consist of one or more polyfunctional polypeptides each of which is made up of modules. Each type I PKS module contains three domains; a β-ketoacyl protein synthase (KS), an acyltransferase (AT) and an acyl carrier protein (ACP). Domains conferring additional enzymatic activities such as ketoreductase (KR), dehydratase (DH) and enoylreductase (ER) can also be found in the PKS modules. These additional domains result in various degrees of reduction of the β-keto groups of the growing polyketide chain. Each module is responsible for one round of condensation and reduction of the β-ketoacyl units. There is a direct correlation between the number of modules and the length of the polyketide chain as well as between the domain composition of the modules and the degree of reduction of the polyketide product. The final polyketide product is released from the PKS protein through the action of a thioesterase domain found in the ultimate module of the PKS system. The genetic organization of most type I PKS enzymes is colinear with the order of biochemical reactions giving rise to the polyketide chain. One skilled in the art will readily understand that these features allow prediction of polyketide core structure based on the architecture of the PKS modules found in a given biosynthetic pathway [Hopwood, *Chem. Rev.*, 97:2465-2497 (1997)].

The compound 2(a) locus PKS system is composed of ORFs 10 to 18 (SEQ ID NOS: 21, 23, 25, 27, 29, 31, 33, 35 and 37) and comprises a total of 27 modules described in Table 3. The first module contains only an ACP domain and corresponds to the loading module (module 0) whereas each of the remaining 26 modules contain domains KS, AT and ACP in various

combinations with KR, DH and ER domains. The thioesterase domain present in ORF 18/module 26 indicates that this module is the ultimate one in the biosynthesis of the polyketide chain. Dehydratase domains in modules 6 and 11 as well as ketoreductase domain in module 12 appear to be inactive due to the presence of non-conservative amino acid residues in highly conserved regions important for catalysis.

Table 3

compound 2(a) locus PKS domain coordinates

ORF no.	SEQ ID NO Amino acid/ Nucleic acid	Amino Acid Residue	Nucleic Acid	Homology	Module no.
10	21/22	57-118	169-354	ACP	0
	21/22	141-566	421-1698	KS	
	21/22	597-1031	1789-3093	AT	1
	21/22	1304-1517	3910-4551	KR	
	21/22	1603-1664	4807-4992	ACP	
	21/22	1690-2118	5068-6354	KS	
	21/22	2135-2562	6403-7686	AT	2
	21/22	2833-3045	8497-9135	KR	
	21/22	3130-3191	9388-9573	ACP	
	21/22	3215-3640	9643-10920	KS	
	21/22	3660-4089	10978-12267	AT	
	21/22	4102-4208	12304-12624	DH	3
	21/22	4612-4829	13834-14487	KR	
	21/22	4911-4972	14731-14916	ACP	
	21/22	5007-5438	15019-16314	KS	
	21/22	5460-5883	16378-17649	AT	4
	21/22	6147-6360	18439-19080	KR	
	21/22	6444-6505	19330-19515	ACP	
	21/22	6529-6954	19585-20862	KS	
	21/22	6979-7402	20935-22206	AT	5
	21/22	7703-7918	23107-23754	KR	
	21/22	8002-8063	24004-24189	ACP	
11	23/24	37-462	109-1386	KS	
	23/24	493-919	1477-2757	AT	
	23/24	932-1038	2794-3114	DH*	6
	23/24	1411-1672	4231-4881	KR	

	23/24	1706-1767	5116-5301	ACP	
	23/24	1794-2215	5380-6645	KS	
	23/24	2232-2659	6694-7977	AT	7
	23/24	2960-3173	8878-9519	KR	
	23/24	3258-3319	9772-9957	ACP	
12	25/26	36-461	106-1383	KS	
	25/26	483-907	1447-2721	AT	
	25/26	919-1027	2755-3081	DH	8
	25/26	1439-1655	4315-4965	KR	
	25/26	1736-1797	5206-5391	ACP	
	25/26	1831-2256	5491-6768	KS	
	25/26	2281-2714	6841-8142	AT	9
	25/26	2981-3194	8941-9582	KR	
	25/26	3287-3339	9832-10017	ACP	
	25/26	3361-3786	10081-11358	KS	
	25/26	3803-4225	11407-12675	AT	
	25/26	4494-4706	13480-14118	KR	10
	25/26	4795-4856	14383-14568	ACP	
	25/26	4880-5304	14638-15912	KS	
	25/26	5323-5748	15967-17244	AT	
	25/26	5761-5866	17278-17598	**DH***	11
	25/26	6294-6510	18880-19530	KR	
	25/26	6599-6660	19795-19980	ACP	
13	27/28	35-460	103-1380	KS	
	27/28	484-920	1450-2760	AT	12
	27/28	1195-1406	3583-4218	**KR***	
	27/28	1490-1551	4468-4653	ACP	
14	29/30	35-460	103-1380	KS	
	29/30	487-918	1459-2754	AT	13
	29/30	1219-1431	3655-4293	KR	
	29/30	1514-1575	4540-4725	ACP	
	29/30	1602-2027	4804-6081	KS	
	29/30	2046-2473	6136-7419	AT	
	29/30	2486-2592	7456-7776	DH	14
	29/30	2980-3196	8938-9588	KR	
	29/30	3287-3339	9832-10017	ACP	
	29/30	3363-3788	10087-11364	KS	
	29/30	3810-4237	11428-12711	AT	
	29/30	4249-4355	12745-13065	DH	15
	29/30	4760-4976	14278-14928	KR	
	29/30	5060-5124	15187-15372	ACP	

15	31/32	35-460	103-1380	KS	
	31/32	480-914	1438-2742	AT	
	31/32	926-1032	2776-3096	DH	16
	31/32	1423-1639	4267-4917	KR	
	31/32	1737-1798	5209-5394	ACP	
	31/32	1822-2247	5464-6741	KS	
	31/32	2263-2690	6787-8070	AT	
	31/32	2703-2809	8107-8427	DH	17
	31/32	3188-3404	9562-10212	KR	
	31/32	3483-3544	10447-10632	ACP	
	31/32	3568-3993	10702-11979	KS	
	31/32	4017-4442	12049-13326	AT	
	31/32	4456-4562	13366-13686	DH	18
	31/32	4978-5194	14932-15582	KR	
	31/32	5285-5346	15853-16038	ACP	
16	33/34	35-460	103-1380	KS	
	33/34	481-917	1441-2751	AT	19
	33/34	1205-1416	3613-4248	KR	
	33/34	1500-1561	4498-4683	ACP	
	33/34	1585-2010	4753-6030	KS	
	33/34	2067-2505	6199-7515	AT	20
	33/34	2786-2998	8356-8994	KR	
	33/34	3083-3144	9247-9432	ACP	
17	35/36	40-465	118-1395	KS	
	35/36	503-941	1507-2823	AT	
	35/36	954-1060	2860-3180	DH	21
	35/36	1456-1672	4366-5016	KR	
	35/36	1751-1812	5251-5436	ACP	
	35/36	1835-2260	5503-6780	KS	
	35/36	2281-2718	6841-8154	AT	
	35/36	2731-2837	8191-8511	DH	22
	35/36	3188-3546	9562-10638	ER	
	35/36	3551-3767	10651-11301	KR	
	35/36	3846-3907	11536-11721	ACP	
	35/36	3932-4357	11794-13071	KS	
	35/36	4373-4803	13117-14409	AT	
	35/36	4815-4921	14443-14763	DH	23
	35/36	5300-5516	15898-16548	KR	
	35/36	5597-5658	16789-16974	ACP	
	35/36	5686-6111	17056-18333	KS	
	35/36	6131-6557	18391-19671	AT	

	35/36	6572-6678	19714-20034	DH	24
	35/36	7062-7288	21184-21834	KR	
	35/36	7363-7424	22087-22272	ACP	
18	37/38	34-459	100-1377	KS	
	37/38	502-926	1504-2778	AT	
	37/38	938-1044	2812-3132	DH	25
	37/38	1420-1636	4258-4908	KR	
	37/38	1715-1776	5143-5328	ACP	
	37/38	1799-2224	5395-6672	KS	
	37/38	2247-2673	6739-8019	AT	
	37/38	2686-2792	8056-8376	DH	26
	37/38	3203-3419	9607-10257	KR	
	37/38	3513-3574	10537-10722	ACP	
	37/38	3649-3872	10945-11616	TE	

One skilled in the art would understand that all KS domains are functional as the multiple amino acid alignment of KS domains present in the compound 2(a) locus PKS system (Figure 2) shows an overall similarity of domains and conservation of amino acid residues and domain regions important for activity. Similarly, multiple amino acid alignment of AT domains (Figure 3), ER domains (Figure 5), ACP domains (Figure 7) and TE domains (Figure 8) show an overall similarity of related domains and a high conservation of protein regions and of amino acid residues important for catalytic activity. The domains that occur only once in the compound 2(a) locus PKS, namely the enoylreductase (ER) domain in ORF 17 (SEQ ID NO: 35) and the thioesterase (TE) domain in ORF 18 (SEQ ID NO: 37) are compared to prototypical domains from the nystatin type I polyketide system (Figures 5 and 8) (see Brauteset *et al., supra*).

Comparison of DH domains found in the compound 2(a) locus PKS indicates a high conservation of amino acid residues important for catalytic activity (Figure 4). However, two DH domains are inactive as they contain non-conservative amino acid substitutions in a region of high sequence conservation. As highlighted in Figure 4, the DH domain of module 6 in ORF 11 (SEQ ID NO: 23) and the DH domain of module 11 in ORF 12 (SEQ ID

NO: 25) contain substitutions of charged amino acids arginine and glutamic acid respectively for non-charged aliphatic amino acids.

Comparison of KR domains found in the compound 2(a) locus PKS system also displays a conservation of active sites and amino acid residues important for catalysis with the exception of the KR domain of module 12 found in ORF 13 (SEQ ID NO: 27). Figure 6 shows the presence in that module of a substitution of a glutamine (Q) for a highly conserved tyrosine (Y) amino acid residue. This non-conservative amino acid substitution results in the inactivation of the enzymatic activity of the KR domain of module 12 in ORF 13 (SEQ ID NO: 27) (ORF13_pKR01).

Phylogenetic analysis of the compound 2(a) locus PKS AT domains was conducted to assess the nature of the β-keto acyl units that are incorporated in the growing polyketide chain. The compound 2(a) locus PKS AT domains were compared to two domains, AAF71779mod03 and AAF71766mod11, derived from the nystatin PKS system [Brautaset, *supra*] and specifying the incorporation of malonyl-CoA and methylmalonyl-CoA respectively. Figure 9 shows the phylogenetic relatedness of the various AT domains indicating that, in the compound 2(a) locus PKS, ORF 13 (SEQ ID NO: 27) module 12 as well as ORF 16 (SEQ ID NO: 33) modules 19 and 20 incorporate methylmalonate in the polyketide chain whereas all remaining AT domains incorporate malonate extender β-keto acyl units.

Domain analysis of the compound 2(a) locus PKS system provides clear indication as to synthesis of the polyketide core structure. While not intending to be limited to any particular mode of action or biosynthetic scheme, the nature and organization of the compound 2(a) locus PKS modules can explain the synthesis of Compound 2(a). Figure 10 highlights schematically a series of reactions catalyzed by the polyketide synthase system based on the correlation between the deduced domain architecture and the polyketide core of the compounds 2(a). Type I PKS domains and the reactions they carry out are well known to those skilled in the art and well documented in the literature; see for example, Hopwood, *supra*.

A biosynthetic pathway for the production of the γ-aminobutyryl-CoA starter unit is also shown. The gene product of ORF 28 (SEQ ID NO: 57), a

member of protein family TMOA, catalyzes the decarboxylative oxidation of arginine forming 4-guanidinobutanamide. The gene product of ORF 33 (SEQ ID NO: 67), a member of protein family ADHY, catalyzes hydrolysis of the amidino group forming γ-aminobutanamide that is further activated by either ORF 27 or 36 (SEQ ID NOS: 55 and 73 respectively), both members of protein family CALB, to give γ-aminobutyryl-CoA (Figure 10a). The gene product of ORF 19 (SEQ ID NO: SEQ ID NO: 39), a member of protein family AYTF, loads this unusual extender unit onto the ACP domain of the loading module (module 0) of ORF 10 (SEQ ID NO: 21), a member of protein family PKSH, as illustrated in Figure 10b. The polyketide chain continues to grow by the sequential condensation of malonyl-CoA and methylmalonyl-CoA extender units that are further reduced by specific domains to various degrees. Dehydratase domains found in module 6 of ORF 11 (SEQ ID NO: 23) and module 11 of ORF 12 (SEQ ID NO: 25) as well as the ketoreductase domain found in module 12 of ORF 13 (SEQ ID NO: 27) are inactive and consequently do not catalyze their respective reductive reactions. The mature polyketide chain is then released through the action of the thioesterase domain found in module 26 of ORF 18 (SEQ ID NO: 37), a member of protein family PKSH as illustrated in Figure 10b. The polyketide core structure expected from the architecture of the PKS domains of the compound 2(a) locus is entirely consistent with the polyketide portion of the compound 2(a).

The compound 2(a) locus contains genes involved in the synthesis of two other components found in the chemical structure of the compound 2(a) locus. Figure 11a illustrates a biosynthetic pathway for the production of the aminohydroxy-cyclopentenone moiety found in the compound 2(a) locus. The gene product of ORF 35 (SEQ ID NO: 71), a member of protein family AYTP, condenses glycine with succinyl-CoA forming 5-aminolevulinate. This intermediate is further activated through the action of either the gene products of ORF 27 or 36 (SEQ ID NOS: 55 and 73 respectively), both members of protein family CALB, forming 5-aminolevulinate-CoA that may spontaneously cyclize to produce aminohydroxycyclopentenone. This moiety is subsequently condensed to the activated carboxy terminus of the polyketide chain through

the action of the gene product of ORF 34 (SEQ ID NO: 69), a member of protein family ADSN as illustrated in Figure 10c.

Figure 11b depicts the biosynthetic pathway of the deoxysugar component of Compound 2(a). The gene product of ORF 24 (SEQ ID NO: 49), a member of protein family NUTA, activates D-glucose forming dNDP-D-glucose that is subsequently dehydrated through the action of the gene product of ORF 25 (SEQ ID NO: 51), a member of protein family DEPA, forming dNDP-4-keto-4, 6-dideoxy-D-glucose. The gene product of ORF 22 (SEQ ID NO: 45), a member of protein family DEPL, further reduces this intermediate forming dNDP-D-fucose that is subsequently epimerized by the gene product of ORF 23 (SEQ ID NO: 47), a member of protein family EPIM, producing dNDP-L-rhamnose.

The final deoxysugar moiety is transferred onto a hydroxyl group of the polyketide core structure through the action of a glycosyltransferase, i.e. the gene product of ORF 9 (SEQ ID NO: 19), a member of protein family GTFA, as illustrated in Figure 10c. Figure 10c proposes one scheme in regard to timing of the reactions catalyzed by the gene product of ORF 34 (SEQ ID NO: 69), a member of protein family CALB, and by the gene product of ORF 9 (SEQ ID NO: 19), a member of protein family GTFA. However, it will be readily understood that the invention does not reside in the actual timing and order of the reactions as depicted in Figure 10c.

Additional proteins forming the compound 2(a) locus include the gene product of ORF 2 (SEQ ID NO: 4) and a member of protein family TESA which is expected to having polyketide-priming editing functions; the gene products of ORFs 3, 4, 5 and 31 (SEQ ID NOS: 6, 8, 10 and 63), members of protein families REGD, RREB, SPKK and REGD respectively, are expected to regulate synthesis of Compound 2(a); the gene products of ORFs 6 and 21 (SEQ ID NOS: 12 and 43), members of protein families UNEW and ABCD respectively, are involved in transmembrane transport; and the gene product of ORF 29 (SEQ ID NO: 59), a member of protein family PPTF, activates ACP domains through phosphopantetheinylation.

Structural modification of compound of Formula I and Formula II and Compound 2(a) are attained by the genetic modifications of the compound 2(a) locus. Genetic modifications of PKS biosynthetic loci are well known in

the art. The WO 01/34816 patent publication teaches the construction of a library of structural variants of the macrolide polyketide rapamycin derived from the genetic modification of genes in the locus that directs rapamycin synthesis. The genetic modifications taught, include gene inactivation, gene insertion and gene replacement. These modifications, both individually and in combination at different positions within the rapamycin locus, resulted in alteration of polyketide starter units, chain length and hydroxyl sterospecificities in rapamycin. Similarly, McDaniel *et.al.* [Proc Natl Acad Sci USA, 1999, 96:18646-51] generated a library of over 50 derivatives of the macrolide antibiotic erythromycin using a combination of genetic modifications including gene inactivation, macrolide chain length and hydroxyl sterospecificity modifications of the erythromycin biosynthesis genes.

The elucidation of the nucleic acid sequences that encodes the biosynthesis of Compound 2a provides the biological tools to enable one skilled in the art to genetically modify the biosynthetic pathway to generate variants of the Compound 2a. In particular, Type I PKS systems may be manipulated by changing the number of modules, their specificities towards carboxylic acids, and by inactivating or inserting domains with reductive activities (Katz, Chem. Rev. v. 97, 2557-2575, 1997). Thus, the polyketide synthase system of Compound 2(a) may be engineered by modifying, adding, or deleting domains, or replacing them with those taken from other Type I PKS enzymes. Compounds of Formula I may be produced using a modified PKS system created based on the polyketide synthase system for the production of Compound 2a. Preferred modified PKS systems are those wherein a KS, AT, KR, DH or ER domain has been inactivated or deleted.

In one aspect, the invention is directed to preparation of a polyketide of Formula I or II resulting from a modified polyketide synthase system, which modification include deletions, mutagenesis, inactivation or replacement of one or more of the domains of the invention. The modified polyketide synthase system produces compounds of Formula I that may differ from the compound of Formula 2a in size, degree of saturation and oxidation. In another aspect, the invention is directed to compounds of Formula I or II produced by genetic modification of the polyketide synthase system for the compound 2(a) locus.

The compounds of this invention may be formulated into pharmaceutical compositions comprised of compounds of Formula I in combination with a pharmaceutically acceptable carrier.

The compounds of this invention are useful in treating bacterial infections, fungal infections and cancer.

Molecular terms, when used in this application, have their common meaning unless otherwise specified.

The term alkyl refers to a linear or branched hydrocarbon group. Examples of alkyl groups include, without limitation, methyl, ethyl, n-propyl, isopropyl, n-butyl, pentyl, hexyl, heptyl, cyclopentyl, cyclohexyl, cyclohexymethyl, and the like. Alkyl groups may optionally be substituted with one or more substituents selected from acyl, amino, acylamino, acyloxy, carboalkoxy, carboxy, carboxyamido, cyano, halo, hydroxyl, nitro, thio, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, sulfinyl, sulfonyl, oxo, guanidino and formyl.

The term alkenyl refers to a linear, branched or cyclic hydrocarbon group containing at least one carbon-carbon double bond. Examples of alkenyl groups include, without limitation, vinyl, 1-propene-2-yl, 1-butene-4-yl, 2-butene-4-yl, 1-pentene-5-yl and the like. Alkenyl groups may optionally be substituted with one or more substituents selected from acyl, amino, acylamino, acyloxy, carboalkoxy, carboxy, carboxyamido, cyano, halo, hydroxyl, nitro, thio, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, sulfinyl, sulfonyl, formyl, oxo and guanidino. The double bond portion(s) of the unsaturated hydrocarbon chain may be either in the cis or trans configuration.

The term cycloalkyl or cycloalkyl ring refers to a saturated or partially unsaturated carbocyclic ring in a single or fused carbocyclic ring system having from three to fifteen ring members. Examples of cycloalkyl groups include, without limitation, cyclopropyl, cyclobutyl, cyclohexyl, and cycloheptyl. Cycloalkyl groups may optionally be substituted with one ore more substituents selected from acyl, amino, acylamino, acyloxy, carboalkoxy, carboxy, carboxyamido, cyano, halo, hydroxyl, nitro, thio, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, sulfinyl, sulfonyl and formyl.

The term heterocycloalkyl, heterocyclic or heterocycloalkyl ring refers to a saturated or partially unsaturated ring containing one to four hetero atoms or hetero groups selected from O, N, NH, NR^x, PO_2, S, SO or SO_2 in a single or fused heterocyclic ring system having from three to fifteen ring members. Examples of heterocycloakyl groups include, without limitation, morpholinyl, piperidinyl, and pyrrolidinyl. Heterocycloalkyl groups may optionally be substituted with one or more substituents selected from acyl, amino, acylamino, acyloxy, oxo, thiocarbonyl, imino, carboalkoxy, carboxy, carboxyamido, cyano, halo, hydroxyl, nitro, thio, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, sulfinyl, sulfonyl and formyl.

The term amino acid refers to a natural amino acid, a synthetic amino acid or a synthetic derivative of a natural amino acid. Examples of natural amino acids include, but are not limited to alanine, arginine, asparagine, aspartic acid, cysteine, glutamic acid, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine and valine.

The term halo is defined as a bromine, chlorine, fluorine or iodine atom.

The term aryl or aryl ring refers to an aromatic group comprising a single or fused ring system, having from five to fifteen ring members. Examples of aryl groups include, without limitation, phenyl, naphthyl, biphenyl, terphenyl. Aryl groups may optionally be substituted with one or more substituent group selected from acyl, amino, acylamino, acyloxy, azido, alkythio, carboalkoxy, carboxy, carboxyamido, cyano, halo, hydroxyl, nitro, thio, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, sulfinyl, sulfonyl and formyl.

The term heteroaryl or heteroaryl ring refers to an aromatic group comprising a single or fused ring system, having from five to fifteen ring members and containing at least one hetero atom such as O, N, S, SO and SO_2. Examples of heteroaryl groups include, without limitation, pyridinyl, thiazolyl, thiadiazolyl, isoquinolinyl, pyrazolyl, oxazolyl, oxadiazoyl, triazolyl, and pyrrolyl groups. Heteroaryl groups may optionally be substituted with one or more substituent groups selected from acyl, amino, acylamino, acyloxy, carboalkoxy, carboxy, carboxyamido, cyano, halo, hydroxyl, nitro, thio,

thiocarbonyl, alkyl, alkenyl, alkynyl, cycloalkyl, heterocyclyl, aryl, heteroaryl, alkoxy, aryloxy, sulfinyl, sulfonyl, and formyl.

As used herein, the term "treatment" refers to the application or administration of a therapeutic agent to a patient, or application or administration of a therapeutic agent to an isolated tissue or cell line from a patient, who has a disorder, e.g., a disease or condition, a symptom of disease, or a predisposition toward a disease, with the purpose to cure, heal, alleviate, relieve, alter, remedy, ameliorate, improve, or affect the disease, the symptoms of disease, or the predisposition toward disease.

As used herein, a "pharmaceutical composition" comprises a pharmacologically effective amount of a farnesyl dibenzodiazepinone and a pharmaceutically acceptable carrier. As used herein, "pharmacologically effective amount," "therapeutically effective amount" or simply "effective amount" refers to that amount of a farnesyl dibenzodiazepinone effective to produce the intended pharmacological, therapeutic or preventive result. For example, if a given clinical treatment is considered effective when there is at least a 25% reduction in a measurable parameter associated with a disease or disorder, a therapeutically effective amount of a drug for the treatment of that disease or disorder is the amount necessary to effect at least a 25% reduction in that parameter.

The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent. Such carriers include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. The term specifically excludes cell culture medium. For drugs administered orally, pharmaceutically acceptable carriers include, but are not limited to pharmaceutically acceptable excipients such as inert diluents, disintegrating agents, binding agents, lubricating agents, sweetening agents, flavoring agents, coloring agents and preservatives. Suitable inert diluents include sodium and calcium carbonate, sodium and calcium phosphate, and lactose, while corn starch and alginic acid are suitable disintegrating agents. Binding agents may include starch and gelatin, while the lubricating agent, if present, will generally be magnesium stearate, stearic acid or talc. If desired, the tablets may be coated with a material such as

glyceryl monostearate or glyceryl distearate, to delay absorption in the gastrointestinal tract.

Pharmaceutically acceptable salts include acid addition salts and base addition salts. The nature of the salt is not critical, provided that it is pharmaceutically-acceptable. Without being limited, examples of acid addition salts include hydrochloric, hydrobromic, hydroiodic, nitric, carbonic, sulphuric, phosphoric, formic, acetic, citric, tartaric, succinic, oxalic, malic, glutamic, propionic, glycolic, gluconic, maleic, embonic (pamoic), methanesulfonic, ethanesulfonic, 2-hydroxyethanesulfonic, pantothenic, benzenesulfonic, toluenesulfonic, sulfanilic, mesylic, cyclohexylaminosulfonic, stearic, algenic, β-hydroxybutyric, malonic, galactantic, galacturonic acid and the like. Suitable pharmaceutically-acceptable base addition salts of compounds of the invention include, but are not limited to, metallic salts made from aluminium, calcium, lithium, magnesium, potassium, sodium and zinc or organic salts made from N,N'-dibenzylethylenediamine, chloroprocaine, choline, diethanolamine, ethylenediamine, N-methylglucamine, lysine, procaine and the like. Additional examples of pharmaceutically acceptable salts are listed in *Journal of Pharmaceutical Sciences*, 1977, 66:2. All of these salts may be prepared by conventional means form the corresponding compounds of Formula I by treating with the appropriate acid or base.

The compounds of the present invention can possess one or more asymetric carbon atoms and can exist as optical isomers forming mixtures of racemic or non-racemic compounds. The compounds of the present invention are useful as a single isomer or as a mixture of stereochemical isomeric forms. Diastereoisomers, i.e., nonsuperimposable stereochemical isomers, can be seperated by conventional means such as chromatography, distillation, crystallization and sublimation. The optical isomers can be obtained by resolution of the racemic mixtures according to conventional processes.

The invention embraces isolated compounds. An isolated compound refers to a compound which represents at least 10%, 20%, 50% and 80% of the compound of the present invention present in a mixture, provided that the mixture comprising the compound of the invention has demonstrable (i.e. statistically significant) biological activity including antibacterial, antifungal or

anticancer activity when tested in conventional biological assays known to a person skilled in the art.

The compounds of the present invention, or pharmaceutically acceptable salts thereof, can be formulated for oral, intravenous, intramuscular, subcutaneous, topical or parenteral administration for the therapeutic or prophylactic treatment of diseases, particularly bacterial and fungal infections. For oral or parental administration, compounds of the present invention can be mixed with conventional pharmaceutical carriers and excipients and used in the form of tablets, capsules, elixirs, suspensions, syrups, wafers and the like. The compositions comprising a compound of this present invention will contain from about 0.1% to about 99.9%, about 5% to about 95%, about 10% to about 80% or about 15% to about 60% by weight of the active compound.

The pharmaceutical preparations disclosed herein are prepared in accordance with standard procedures and are administered at dosages that are selected to reduce, prevent, or eliminate bacterial and fungal infection or the cancer (See, e.g., Remington's Pharmaceutical Sciences, Mack Publishing Company, Easton, PA and Goodman and Gilman's the Pharmaceutical Basis of Therapeutics, Pergamon Press, New York, NY, the contents of which are incorporated herein by reference, for a general description of the methods for administering various antimicrobial agents for human therapy). The compositions of the present invention can be delivered using controlled (e.g., capsules) or sustained release delivery systems (e.g., bioerodable matrices). Exemplary delayed release delivery systems for drug delivery that are suitable for administration of the compositions of the invention (preferably of Formula I) are described in U.S. Patent Nos 4,452,775 (issued to Kent), 5,239,660 (issued to Leonard), 3,854,480 (issued to Zaffaroni).

The pharmaceutically-acceptable compositions of the present invention comprise one or more compounds of the present invention in association with one or more non-toxic, pharmaceutically-acceptable carriers and/or diluents and/or adjuvants and/or excipients, collectively referred to herein as "carrier" materials, and if desired other active ingredients. The compositions may contain common carriers and excipients, such as corn starch or gelatin,

lactose, sucrose, microcrystalline cellulose, kaolin, mannitol, dicalcium phosphate, sodium chloride and alginic acid. The compositions may contain crosarmellose sodium, microcrystalline cellulose, sodium starch glycolate and alginic acid.

Lubricants that can be used include magnesium stearate or other metallic stearates, stearic acid, silicon fluid, talc, waxes, oils and colloical silica.

Flavouring agents such as peppermint, oil of wintergreen, cherry flavouring or the like can also be used. It may also be desirable to add a coloring agent to make the dosage form more esthetic in appearance or to help identify the product comprising a compound of the present invention.

For oral administration, the pharmaceutical compositions are in the form of, for example, a tablet, capsule, suspension or liquid. For oral use, solid formulations such as tablets and capsules are particularly useful. Sustained released or enterically coated preparations may also be devised. Tablet binders that can be included are acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Providone), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose. For pediatric and geriatric applications, suspension, syrups and chewable tablets are especially suitable. The pharmaceutical composition is preferably made in the form of a dosage unit containing a therapeutically-effective amount of the active ingredient. Examples of such dosage units are tablets and capsules. For therapeutic purposes, the tablets and capsules can contain, in addition to the active ingredient, conventional carriers such as binding agents, for example, acacia gum, gelatin, polyvinylpyrrolidone, sorbitol, or tragacanth; fillers, for example, calcium phosphate, glycine, lactose, maize-starch, sorbitol, or sucrose; lubricants, for example, magnesium stearate, polyethylene glycol, silica or talc: disintegrants, for example, potato starch, flavoring or coloring agents, or acceptable wetting agents. Oral liquid preparations generally are in the form of aqueous or oily solutions, suspensions, emulsions, syrups or elixirs may contain conventional additives such as suspending agents, emulsifying agents, non-aqueous agents, preservatives, coloring agents and flavoring agents. Examples of additives for liquid preparations include acacia, almond

oil, ethyl alcohol, fractionated coconut oil, gelatin, glucose syrup, glycerin, hydrogenated edible fats, lecithin, methyl cellulose, methyl or propyl *para*-hydroxybenzoate, propylene glycol, sorbitol, or sorbic acid.

For intravenous (IV) use, compounds of the present invention can be dissolved or suspended in any of the commonly used intravenous fluids and administered by infusion. Intravenous fluids include, without limitation, physiological saline or Ringer's solution.

Formulations for parental administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions. These solutions or suspensions can be prepared from sterile powders or granules having one or more of the carriers mentioned for use in the formulations for oral administration. The compounds can be dissolved in polyethylene glycol, propylene glycol, ethanol, corn oil, benzyl alcohol, sodium chloride, and/or various buffers.

For intramuscular preparations, a sterile formulation of compounds of the present invention or suitable soluble salts forming the compound, can be dissolved and administered in a pharmaceutical diluent such as Water-for-Injection (WFI), physiological saline or 5% glucose. A suitable insoluble form of the compound may be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, e.g. an ester of a long chain fatty acid such as ethyl oleate.

For topical use the compounds of present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of creams, ointments, liquid sprays or inhalants, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Alternatively, the compound of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery. In another embodiment, the unit dosage form of the compound can be a solution of the compound or a salt thereof in a suitable diluent in sterile, hermetically sealed ampoules.

The amount of the compound of the present invention in a unit dosage comprises a therapeutically-effective amount of at least one active compound of the present invention which may vary depending on the recipient subject, route and frequency of administration. A recipient subject refers to a plant, a cell culture or an animal such as an ovine or a mammal including a human.

According to this aspect of the present invention, the novel compositions disclosed herein are placed in a pharmaceutically acceptable carrier and are delivered to a recipient subject (including a human subject) in accordance with known methods of drug delivery. In general, the methods of the invention for delivering the compositions of the invention in vivo utilize art-recognized protocols for delivering the agent with the only substantial procedural modification being the substitution of the compounds of the present invention for the drugs in the art-recognized protocols.

Likewise, the methods for using the claimed composition for treating cells in culture, for example, to eliminate or reduce the level of bacterial or fungal contamination of a cell culture, utilize art-recognized protocols for treating cell cultures with antibacterial or antifungal agent(s) with the only substantial procedural modification being the substitution of the compounds of the present invention for the agents used in the art-recognized protocols.

The compounds of the present invention provide a method for treating bacterial infections, fungal infections and pre-cancerous or cancerous conditions. As used herein the term unit dosage refers to a quantity of a therapeutically-effective amount of a compound of the present invention that elicits a desired therapeutic response. As used herein the phrase therapeutically-effective amount means an amount of a compound of the present invention that prevents the onset, alleviates the symptoms, or stops the progression of a bacterial infection, fungal infection or pre-cancerous or cancerous condition. The term treating is defined as administering, to a subject, a therapeutically-effective amount of at least one compound of the

present invention, both to prevent the occurrence of a bacterial or fungal infection or pre-cancer or cancer condition, or to control or eliminate a bacterial or fungal infection or pre-cancer or cancer condition. The term desired therapeutic response refers to treating a recipient subject with a compound of the present invention such that a bacterial or fungal infection or pre-cancer or cancer condition is reversed, arrested or prevented in a recipient subject.

The compounds of the present invention can be administered as a single daily dose or in multiple doses per day. The treatment regime may require administration over extended periods of time, e.g., for several days or for from two to four weeks. The amount per administered dose or the total amount administered will depend on such factors as the nature and severity of the infection, the age and general health of the recipient subject, the tolerance of the recipient subject to the compound and the type of the bacterial or fungal infection, or type of cancer.

A compound according to this invention may also be administered in the diet or feed of a patient or animal. The diet for animals can be normal foodstuffs to which the compound can be added or it can be added to a premix.

The compounds of the present invention may be taken in combination, together or separately with any known clinically approved antibiotic, anti-fungal or anti-cancer to treat a recipient subject in need of such treatment.

Compounds of Formula I are obtained biosynthetically by culturing *Actinomycetes* species in growth media described in Table 4, at temperatures between 24^0 C – 34^0 C and with shaking to aerate of the culture medium for 3 to 40 days. The compounds of Formula I are extracted and isolated from the bacterial culture by methods known to a skilled person including centrifugation, chromatography, adsorption, filtration, extraction or other methods of separation.

The compounds of Formula I may be biosynthesized by various microorganisms. Microorganisms that may synthesize the compounds of the present invention include but are not limited to bacteria of the order Actinomycetales, also referred to as actinomycetes. Non-limiting examples of members belonging to the genera of Actinomycetes include *Nocardia,*

Geodermatophilus, Actinoplanes, Micromonospora, Nocardioides, Saccharothrix, Amycolatopsis, Kutzneria, Saccharomonospora, Saccharopolyspora, Kitasatospora, Streptomyces, Microbispora, Streptosporangium, Actinomadura. The taxonomy of actinomycetes is complex and reference is made to Goodfellow (1989) Suprageneric classification of actinomycetes, *Bergey's Manual of Systematic Bacteriology*, Vol. 4, Williams and Wilkins, Baltimore, pp 2322-2339, and to Embley and Stackebrandt, (1994), and *The molecular phylogeny and systematics of the actinomycetes, Annu. Rev. Microbiol.* 48, 257-289 (1994), for genera that may synthesize the compounds of the invention, incorporated herein in their entirety by reference.

Microorganisms biosynthetically producing compounds of Formula I are cultivated in culture media containing known nutritional sources for actinomycetes having assimilable sources of carbon, nitrogen plus optional inorganic salts and other known growth factors at a pH of about 6 to about 9, non-limiting examples of growth media are provided in Table 4 below. Microorganisms are cultivated at incubation temperatures of about 20^0 C to about 40^0 C for about 3 to about 40 days.

Table 4. Examples of Growth Media for Production of Compounds of Formula I

Component	VA	QB	GA*4	MA	NA	KH	OA	HA	RM	EA	KA	CA
pH*5	7	7.2		7.5	7	7	7		6.85	7	5.7	7
Glucose	50	12	10			10	10	10	10	5	10	10
Sucrose			103					340	100			
Lactose										50		
Cane molasses					10							15
Soluble starch		10		25								
Potato dextrin						20						40
Corn steep										5		
Corn steep		5					3				10	
Dried yeast				2							5	
Yeast extract			5			5	3	3	5			
Malt extract							3	3				
Pharmamedia™		10										
Glycerol					20		5			15	5	
NA-Amine A						5						10
Soybean				15							10	
Soybean flour	30									10		
Beef extract							3					
Bacto-peptone					1			5		5		
$MgSO_4.7H_2O$										0.5		1
$MgCl_2.6H_2O$			10.12									
$CaCO_3$	6			4	4	1	2			3	2	2
NaCl	5			5							5	
$(NH_4)_2SO_4$	3			2						2		
K_2SO_4			0.25						0.25			
$MnCl_2.4H_2O$										0.1		
$MgCl_2.6H_2O$								1	10			
$FeCl_2.4H_2O$										0.1		
$ZnCl_2$										0.1		
Thiamine							0.1					
Casamino acid			0.1		5				0.1			
Proflo oil		4										
MOPS									21			
Trace element solution *3 ml/L									2			

Unless otherwise indicated all the ingredients are in gm/L.

*3 Trace elements solution contains: $ZnCl_2$ 40 mg; $FeCl_3 6H_2O$ (200 mg); $CuCl_2 2H_2O$ (10 mg); $MnCl_2.4H_2O$; $Na_2B_4O_7.10H_2O$ (10mg); $(NH_4)_6Mo_7O_{24}.4H_2O$ (10 mg) per litre.

*4 Dissolve components in 800 ml water and autoclave, later add: 10 ml KH_2PO_4 (0.5% solution); 80 ml $CaCl_2.2H_2O$ (3.68 % solution); 15 ml L-proline (20% solution); 100 ml TES buffer (5.73% solution, pH 7.2); 5 ml NaOH (1N solution), and 2 ml of trace elements solution.

*5 The pH is to be adjusted as marked prior to the addition of $CaCO_3$ in those media containing it.

The culture media inoculated with the microorganisms which biosynthetically produce compounds of Formula I, may be aerated by incubating the inoculated culture media with agitation, for example shaking on a rotary shaker, or a shaking water bath. Aeration may also be achieved by the injection of air, oxygen or an appropriate gaseous mixture to the inoculated culture media during incubation.

After cultivation and production of compounds of Formula I, the compounds can be extracted and isolated from the cultivated culture media by techniques known to a skilled person in the art and/or disclosed herein, including for example centrifugation, chromatography, adsorption. For example, the cultivated culture media can be mixed with a suitable organic solvent such as n-butanol, n-butyl acetate and 4-methyl-2-pentanone, the organic layer can be separated for example, by centrifugation followed by the removal of the solvent, by evaporation to dryness or by evaporation to dryness under vacuum. The resulting residue can optionally be reconstituted with for example water, ethanol, ethyl acetate, methanol or a mixture thereof, and re-extracted with a suitable organic solvent such as hexane, carbon tetrachloride, methylene chloride or a mixture thereof. After removal of the solvent, the compound of Formula I can be further purified by the use of standard techniques such as chromatography.

The compounds of Formula I that are biosynthesized by microorganisms may optionally be subjected to random and/or directed chemical modifications to form compounds that are derivatives or structural analogs of compounds of Formula I. Derivatives or structural analogs of compounds of Formula I having similar functional activities are within the scope of the present invention. Compounds of Formula I may optionally be modified using methods known in the art and described herein.

Unless otherwise indicated, all numbers expressing quantities of ingredients and properties such as molecular weight, reaction conditions, IC_{50} and so forth used in the specification and claims are to be understood as being modified in all instances by the term "about". Accordingly, unless indicated to the contrary, the numerical parameters set forth in the present specification and attached claims are approximations. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the

scope of the claims, each numerical parameter should at least be construed in light of the number of significant figures and by applying ordinary rounding techniques. Notwithstanding that the numerical ranges and parameters setting forth the broad scope of the invention are approximations, the numerical values set in the examples, Tables and Figures are reported as precisely as possible. Any numerical values may inherently contain certain errors resulting from variations in experiments, testing measurements, statistical analyses and such.

The compounds of Formula I, Formula II and compound 2(a) may optionally be chemically modified using methods known in the art and described herein.

The compounds of the invention are made by biofermentation and well-known chemical schemes. The schemes described herein are exemplary, any chemical synthetic process known to a person skilled in the art providing the structures described herein, may be used and are therefore comprised in the present invention.

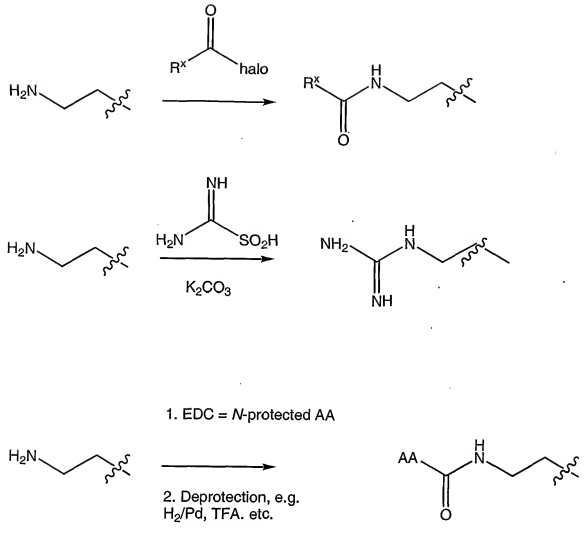
SCHEME 1 Acylation Reactions

EDC = 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide

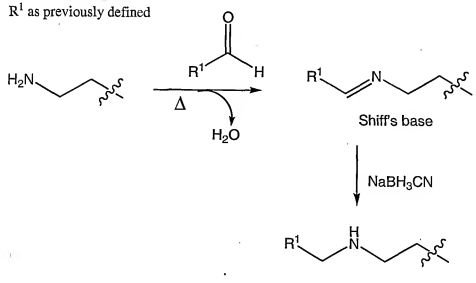
Protective groups include N-benzyloxycarbonyl (CBZ), N-butoxycarbonyl (BOC), N-fluoren-9-ylmethoxycarbonyl (FMOC)

R^x represents C_{1-6} alkyl, C_{2-6} alkenyl, aryl or heteroaryl

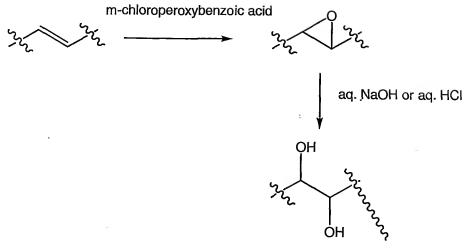
AA represents a naturally occurring amino acid



Scheme 2. Aminations/reductive aminations of terminal nitrogen

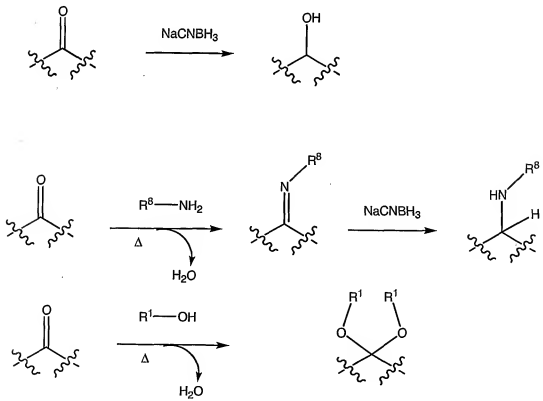


Scheme 3. Olefin reactions



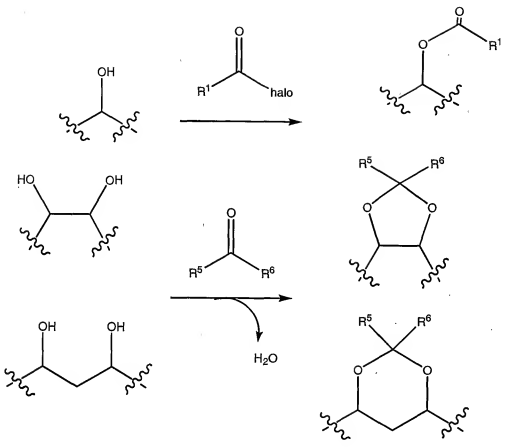
Scheme 4. Ketone reactions

R^1 and R^8 are as previously defined.

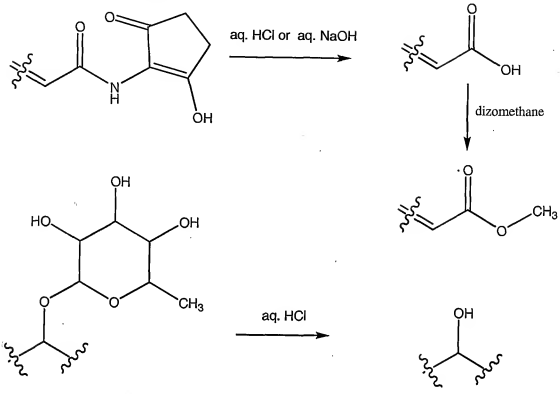


Scheme 5. *O*- Reactions

R^1, R^5 and R^6 are as peviously defined.



Scheme 6. Hydrolysis/Esterification



Scheme 1 is used to obtain Compounds 2(m), 2(n), 2(o), 2(p), 2(q), 2(r), 2(s), 2(t), 2(u), 2(v), 2(w), 2(x), 2(y), 2(z), 2(aa), and 2(ab) from Compound 2(a).

Scheme 3 is used to obtain Compound 2(b) from Compound 2(a).

Scheme 4 is used to obtain Compounds 2(c), 2(d), 2(e) and 2(f) from Compound 2(a).

Scheme 6 is used to obtain Compounds 2(g), 2(h), 2(i) and 2(j) from Compound 2(a).

The features of the invention are further described below by way of examples and are not to be construed as limiting in their scope.

Example 1 Production of Compound 2(a) by Fermentation

Example 1(A): Preparation of Strain [C03U03]023

Strain [C03]023: *Streptomyces aizunensis* NRRL B-11277 was plated on three tomato paste oatmeal agar (ATCC medium 1360) plates for sporulation at 28 °C. The plates were incubated for a period of 5-7 days, after which spores were collected from each plate into 5 ml sterile distilled water, spun down by centrifugation at 5000 rpm (10 min), and dispersed in 20 ml sterile water. After a second centrifugation under the same conditions the pellet was resuspended in 10 ml sterile distilled water. A series of ten-fold dilutions of the original spore suspension were prepared and 0.5 ml aliquots plated on tomato paste-oat meal agar until sporulation occurred (5-7 days). Each individual clone from the plates with single well-isolated colonies (generated from 10^{-8} to 10^{-10} dilutions of the spore suspension) was chosen and transferred to one plate of tomato paste-oat meal agar to generate spores for storage. Each clone was grown in 25x150 mm glass tubes for its production of Compound 2(a). A total of 385 clones were tested for production levels of Compound 2(a). Clone [C03]023 showed a production of 3 times better than the wild-type strain. This clone was chosen, stored, and used for mutagenesis.

Strain [C03U03]023: An aqueous spore suspension of [C03]023 was mutagenized by UV radiation (254 nm) at different energy levels (expressed as mJoules per surface area). Clone [C03U03]023 obtained at 0.4 mJ/1 cm^2 showed slightly more than three times better production than the parent clone [C03]023. Production of Compound 2(a) by the new clone has been consistently reproducible both in shaken flask (500 ml medium QB or VA in 2-L baffled flasks) and in 100-L fermentors with medium VA.

Example 1(B) Activation of lyophilized sample of Strain [C03U03]023

Strain [C03U03]023 was provided as a lyophilized pellet. The lyophilized sample was opened under aseptic conditions, and 0.3-0.5 ml of medium ITSB was added to the sample to make a cell suspension. The cell suspension was transferred to 25 ml of medium ITSB (described below) in a 125-ml flask to form a liquid culture. The liquid culture was incubated at 28 °C

for 3-5 days until visible growth occurred. Purity of the culture was tested by streaking a loop on ISP2 agar plate.

Example 1(C): Preparation and Storage of glycerol stocks of Strain [C03U03]023

Strain [C03U03]023 was grown for 7-10 days at 28°C on several tomato paste-oat meal agar plates. Surface growth was collected from each plate into 5 ml sterile distilled water, spun down by centrifugation at 5000 rpm (10 min), and dispersed in 10 ml sterile water. After a second centrifugation under the same conditions the pellet was resuspended in 2 ml sterile 25% glycerol and 0.5-ml aliquots were stored at –80 °C in screw-capped vials. In addition to the glycerol stocks, the collected cell mass could be resuspended in 15% sterile skim milk and dispensed in 0.5-ml aliquots into glass ampoules and lyophilized following standard procedures.

Example 1(D): Preparation of Seed Culture

A vial containing frozen mycelia prepared as described in Example 1(C) was taken out of freezer and kept on dry ice. Under aseptic conditions, a loopfull of the frozen culture was taken and streaked on the surface of tomato paste-oat meal agar plate and incubated at 28°C until vegetative mycelium appeared (5-7 days). In order to start the seed culture, 2-3 loopfull of the surface growth obtained from the tomato paste-oat meal agar plate was transferred to a 1.5-ml Eppendorf tube containing 300 µl of medium ITSB. The mycelium with agar fragments was homogenized, and 1 ml of medium ITSB was added to the suspension. The content was used to inoculate two 125-ml flasks containing 25 ml of sterile medium ITSB. The flasks were incubated at 28°C for 65-70 hours in a rotary shaker at 250 rpm. This seed culture was then used to inoculate production medium QB or VA.

Example 1(E): Production of Compound 2(a) by Fermentation

A sample of the seed culture prepared as described in Example 1(D) above was checked microscopically for any possible contamination. A sample of the seed culture was then streaked onto one ISP2 plate (control plate) and incubated at 28 °C. From the seed culture under aseptic conditions, 10 ml was

taken and used to inoculate each 2 Liter baffled flask containing 500 ml of sterile medium QB or VA. The fermentation batches were incubated aerobically with shaking (250 rpm) at 28°C for a period of 7 days. After 3-5 days of incubation the control plate was checked for purity of the culture.

The compositions of the growth media used in Examples 1(A) – 1(E) are given below. Note that either of Production media QB or VA may be used in the production of Compound 2(a); however, production medium VA is preferred when conducting the fermentation on a large scale.

Seed Medium ITSB:

Trypticase Soy Broth	(Difco)	30 g
Yeast extract	(Sigma)	3 g
$MgSO_4$	(Sigma)	2 g
Glucose	(Sigma)	5 g
Maltose	(Sigma)	4 g
Distilled water		1 L

Production Medium VA

Glucose	50g
Soybean Flour	30g
$CaCO_3$	6g
NaCl	5g
$(NH_4)_2SO_4$	3g
Distilled water	1L

Production Medium QB:

Soluble starch	(Sigma)	10 g
Glucose	(Sigma)	12 g
Pharmamedia	(Traders protein)	10 g
Corn steep liquor	(Sigma)	5 g
Proflo oil	(Traders Protein)	4 mL *
Distilled water		1 L

* Adjust pH to 7.2, then add Proflo oil

Tomato paste Oatmeal Agar:

Baby Oatmeal Food (Heinz)	20 g
Tomato Paste	20 g
Agar	15 g
Tap water	1 L
pH 7.0	

The production of Compound 2(a) may also be carried out in the production media having the compositions as indicated in Table 4, *supra,* in order of preference.

Example 2 Isolation of Compound 2(a)

Thirty minutes prior to harvest of Compound 2(a) from the fermentation broth of the baffled flasks of Example 1E, regenerated, water-washed, Diaion HP-20® in a quantity of wet-packed volume equal to 12% of the initial fermentation beer volume was added to the whole fermentation broth of Example 1E and modest agitation was continued for 30 minutes. At harvest the fermentation broth from 2 x 500 ml flasks was centrifuged and the supernatant was decanted from the resin and mycelia pellet. The pellet was resuspended in 15% MeOH in water (half the original fermentation beer volume), agitated mildly and recentrifuged, and the surpernatant was decanted from the residue. The residue was washed a second time in the same manner with another 15% MeOH in water, followed by a single final wash with methanol: water (7:3 v/v) (half the original fermentation beer volume) to obtain a well-washed residue. The well-washed mycelia:resin residue was extracted three times with 100% ethanol, each extract being at 20% original beer volume. The three extracts were combined and concentrated under vacuum on a rotary evaporator, to dryness.

The three extracts (representing material from 2 x 500 ml flasks) were combined, filtered on paper and concentrated under vacuo to remove organic solvents. The resulting semi-solid residue (aqueous suspension) of crude Compound 2(a) represented greater than 90% of the respective compounds produced and was about 25% pure. The aqueous suspension was freeze-dried overnight to give 460 mg of a dark brown solid. The solid was stirred

with 10 ml of methanol and centrifuged for 2 minutes to remove insoluble matter.

The semi-solid residue of crude Compound 2(a) was then purified using a Waters Xterra® preparative MS C-18 column with 10 µm packing of dimensions 19 mm diameter x 150 mm length, using the following gradient table (Table 5) from 5mM aqueous ammonium bicarbonate to acetonitrile.

Table 5:

Time (min)	% Aqueous	% Acetonitrile
0	70	30
5	45	55
10	70	30

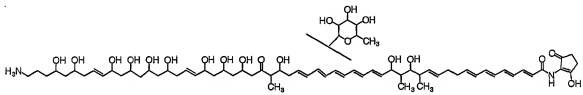
The eluate was monitored at 390 nm, a single run was loaded with 23 mg of crude residue in 0.5 ml of methanol, and a conservative cut of the peak eluting at 3.4 minutes afforded compound 2(a). Nineteen runs were conducted to yield 33 mg of product with about 95% purity.

Example 3 Structural Determination of Compound 2(a)

The structure of compound 2(a) was determined by a combination of genomic information and spectroscopic data, including Mass, UV, and NMR spectroscopy. The Mass was determined by electrospray mass spectrometry to be 1297 (Figure 13) and the UV λ_{max} were found to be 319, 333, 350 (Figure 14). The NMR data were collected at 500 MHz with the compound 2(a) dissolved in MeOH-*d4*, and included proton (Figure 15A), carbon-13 (Figure 15B), and multidimensional pulse sequences gDQCOSY, gHSQC, gHMBC, and TOCSY (Figures 15C, 15 D, 15E and 15F, respectively).

Streptomyces aizunensis NRRL B-11277 was grown on oat meal agar plates for 5-7 days. The surface growth was collected and washed with water, and DNA was extracted following standard procedures (T. Kiesser *et al.* Practical Streptomyces Genetics, The John Innes Foundation, Norwich, UK, 2000). The genomic library was produced in cosmid and plasmid vectors, and the genome was scanned for the presence of gene sequence tags (GSTs) related to the biosynthesis of secondary metabolites as described in E. Zazopoulos et al., Nature Biotechnology 21:187-190 (2003). The GSTs

were used to isolate cosmids containing the compound 2(a) locus. The PKS system found within the compound 2(a) locus was determined to contain 9 PKS genes containing 27 modules. (The analysis of this PKS system is fully described elsewhere herein; see, e.g., Table III and accompanying text). Full analysis of the PKS and associated genes led to the prediction of a structure of Formula 1 below.



1

The position of the glycosidic linkage to the sugar moiety could not be determined by the genomic analysis; however, the positioning of the aminohydroxycyclopentenone unit was determined by analogy with its placement in other actinomycete metabolites (Colabomycin A from *Streptomyces griseoflavus* Tue 2880, J. Antibiot. 1988, 41, 1178-85, 1186-1195 or Enopeptin-A from *Streptomyces griseus*, Osada et al., J. Antibiot. 44, 1463-6 1991).

To obtain expression of these genes, and the end product of this biosynthesis pathway, *S. aizunensis* NRRL B-11277 was grown in several different media designed for the production of secondary metabolites in shaken flasks. At harvest the broth was diluted with an equal volume of methanol to induce cell lysis, and the diluted, clarified broth was concentrated 10 fold. An aliquot (50 µL) from the concentrate from each medium was chromatographed on a Waters Xterra C-18 HPLC column (19 x 150 mm) at a flow rate of 1mL/min and monitored by diode array detector (DAD) UV and positive and negative ion MS. Fractions (800 µL) were collected and tested for antimicrobial activity against a panel of indicator strains. From the extracts of several different media, HPLC fractions in the number 39 to 45 region exhibited strong activity against *Candida albicans* and this correlated with a UV absorption λ_{max} 319, 333, and 351 nm, and with strong MS peaks at m/z 1298 (positive ion mode) and 1296 (negative ion mode). These physical characteristics were entirely consistent with a metabolite of formula 1.

A high yielding medium was chosen and the organism was regrown on a 2-liter scale. The compound 2(a) was extracted from the mycelial pellet with methanol and acetone, and from the broth with Diaion HP-20® resin, from which it was recovered with methanol after the resin had been washed with methanol/water 3:2. The crude extracts were purified by HPLC on a Waters Xterra C-18 column (19 x 150 mm) using an aqueous (5 mM ammonium bicarbonate) / acetonitrile gradient.

Compound 2(a), a yellow solid of MW 1297 Da ($C_{70}H_{108}N_2O_{20}$ requires 1296.75) λ_{max} 319, 334, and 351 nm was the subject of a series of 1D and 2D NMR measurements including a CMR, ^{1}H-NMR, gDQCOSY, gHSQC, gHMBC, TOCSY, gHSQCTOXY, and several 1D TOCSY experiments. See Figures 15A – 15E. Analysis of these spectra led to the assignments shown for compound 2(a) in Figure 17. Although considerable overlap of signals rendered unambiguous assignments of all of the signals to specific protons and carbons impractical, those that could be made unambiguously confirmed the structure predicted from the genomics. A major cross peak in the gHMBC spectrum between the well separated proton resonance at 4.01 ppm and the anomeric carbon at 102.6 ppm placed the sugar as shown, as this proton falls within a 14 carbon section of the major chain with fully assigned carbon and proton signals. A well resolved carbon spectrum with high signal to noise ratio showed that the unassigned methylene carbons were at 42.0, 45.3, 45.4 and 46.6 ppm. Analysis by gHSQC indicates that that these were attached to protons at 2.24, 1.62, 1.50 and 1.68, and 1.55 ppm respectively. Similarly the unassigned carbinols at 66.2, 66.2 (resolved), 67.2 and 69.0 ppm attached to protons at 4.06, 4.08, 4.22 and 3.89 ppm respectively and the unassigned olefinic carbons at 129.1, 131.0, 131.9, 133.3, 133.7, 134.3, 134.8, 136.5, and 138.0 ppm attached to protons at 5.72, 5.72, 6.28, 6.25, 6.28, 6.25, 6.19, 5.53, and 5.86 respectively. The aminohydroxycyclopentenone signals were not straightforward and reflected the tautomeric equilibrium of this moiety. The upfield methylene signal and the downfield carbonyl signals were only 10% of the intensity of those from the other tautomer. The signal from C-1 of this moiety was not detected, a phenomenon which has been previously ascribed to tautomerization for the same structural unit. See, He, H.; Shen, B.;, Korshalla, J.; Siegel, M.M.; Carter,G.T. *J. Antibiot.* 2000, 53, 191~195.

Example 4 Minimal Inhibitory Concentration (MIC) Determination for Compound 2(a)

The MIC determination for fungal and bacterial organisms was performed using the broth microdilution assay adapted from National Committee for Clinical Laboratory Standards (NCCLS) M27-A (Vol. 17 No. 9, 1997), Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts; Approved Standard guidelines: M23-A: Reference Method for Broth Dilution Antifungal Susceptibility Testing of Filamentous Fungi; Approved Standard, vol. 22, No. 16.

Materials:

1) Overnight broth cultures of bacterial and fungal strains to be tested;
2) Stock solution of Compound 2(a) at 3.2 mg/ml in DMSO;
3) Standard 96 well round-bottom plates, sterile;
4) Cation adjusted Mueller-Hinton broth, or Brain Heart Infusion broth (for antibacterial testing);
5) Morpholinepropanesulfonic acid (MOPS)-buffered RPMI-1640 medium (for antifungal testing);
6) Sterile isotonic saline (0.85%);
7) McFarland 0.5 Barium Sulfate Turbidity Standard at 100 X 3.2mg/ml.

Test compound preparation: The test article was prepared as 100x stock solutions in DMSO, with concentrations ranging from 3.2 mg/ml to 0.0625 mg/ml (a two-fold dilution series over 10 points). The first dilution (3.2mg/ml) was prepared by resuspending 0.5 mg of each test article in 156.25 µl of DMSO. The stock is then serially diluted by two-fold increments to obtain the desired concentration range.

Inoculum preparation: For fungal strains, the inoculum was prepared as follows. From an overnight culture in Yeast Media broth, cell density was

adjusted in 0.85% saline to 0.5 McFarland. This procedure yielded a stock suspension of about 5 X 10^6 cells/ml. Following thorough vortexing, a working suspension was prepared by diluting the stock 1:50 in RPMI 1640, and then further diluting it 1:20 with RPMI 1640 to obtain the 2x test inoculum (about 5 X 10^3 cells/ml). For filamentous fungi, the inoculum was prepared as follows. From a spore suspension kept at 4°C, an appropriate dilution in 0.85% saline was made to obtain a final optical density 600 between 0.09-0.11. A working suspension was then prepared by diluting the spore suspension 50 times in RPMI to obtain the 2x test inoculum (about 1 X 105 CFU/ml).

MIC Determination: The 100X test article solutions were diluted 50 times in RPMI 1640, MH or BHI media and dispensed in a 96 well plate, one concentration per column, 10 columns in total. The 11th column contained RPMI 1640 with 1% DMSO with cells, the 12th column contained 100 µl of RPMI 1640 alone.

50 µl of the final cell dilution (yeast, filamentous fungi or bacteria) of each indicator strain was added to each corresponding well of the microplate containing 50 µl of diluted drug or media alone. Assay plates were incubated at 35°C for up to 72 hrs. MIC readings were determined at 24 and 48 hrs for the *Candida* and *Aspergillus* species, and at 48 and 72 hrs for *Cryptococcus neoformans*. MIC readout for each indicator was determined as the lowest concentration of test compound resulting in total absence of growth.

Table 6: MIC (µg/ml) for Compound 2(a) for various strains of yeast and fungi

	MIC (µg/ml)	
Yeasts and filamentous fungi	24 hrs	48 hrs
Candida albicans **ATCC 10231**	4	4
Candida krusei **LSPQ 0309**	8	8
Candida glabrata **LSPQ 0250**	4	8
Candida lusitaniae **ATCC 200953**	4	4
Saccharomyces cerevisiae **ATCC 9763**	4	4
Cryptococcus neoformans **ATCC 32045**	2*	4**
Aspergillus flavus **ATCC 204304**	4	8
Aspergillus fumigatus **ATCC 204305**	16	16

* 48 hrs reading; ** 72 hrs reading

Example 5. In vitro activity of compound 2(a) against *Aspergillus species*

To determine the antifungal activity of compound 2(a) against *Aspergillus* species (*A. fumigatus* and *A. flavus*) a disk diffusion assay was used to determine the minimum effective concentration (MEC) as described by Wong GK, Griffith S, Kojima I and Demain AL. Antifungal activities of rapamycin and its derivatives, prolylrapamycin, 32-desmethylrapamycin, and 32-desmethoxyrapamycin. J. Antibiotics, 51(5): 487-491,1998. Such assay is commonly used to reveal activity of antifungal drugs against filamentous fungi such as *Aspergillus sp.* (Arikan S, Yurdakul P, Hascelik G. Comparison of two methods and three end points in determination of in vitro activity of micafungin against *Aspergillus spp. Antimicrobial Agents and Chemotherapy 47(8): 2640-2643*, 2003).

Preparation of the inoculum: After spreading on YM agar (in cell culture flasks), *Aspergillus* strains (*A. flavus* – ATCC 204304 and *A. fumigatus* - LSPQ 204305) were left sporulating for 4 to 5 days at 35°C. After the addition

of 10 to 20 ml of saline solution (0.85% NaCl), spores were collected by gently rubbing the surface of the conidiophores with a disposable inoculation loop. *Aspergillus* spore suspensions, kept at 4°C, were used as the inoculum for the disc assays.

Preparation of the disks : Stock solutions (5 mg/ml) in methanol and dilutions (0.25, 0.5, 1.0, 2.5, 5.0, 7.5, 10.0 and 50.0 µg/ml), prepared by serial dilutions of stock solution in methanol were prepared for the test article and each of the control compounds. Itraconazole and casponfungin were used as positive controls while fluconazole or DMSO alone were used as negative controls. Drug-containing disks were prepared by spotting of 10 µl of the proper drug solution (or methanol as control) onto filter disks that were then allowed to air-dry.

Agar plate preparation: *Aspergillus* spore suspensions were adjusted to about 81% of transmittance at 530 nm in saline solution. 200 µl of the adjusted inoculum was then mixed with 50 ml of melted 0.8% YM agar (cooled to ~50°C), mixed thoroughly and poured in a 150 mm Petri dish. Once the agar was set, the prepared filters were loaded onto the plates, which were incubated at 35°C. The zone of inhibition (ZOI) of fungal growth was measured after 24 hours of incubation.

Results: Data presented in Table 7 show the lowest concentration (MEC) inducing inhibition of the fungal growth and the corresponding ZOI obtained at this concentration for compound 2(a) and the controls. Results demonstrated that compound 2(a) was active against *Aspergillus fumigatus* and *Aspergillus flavus*. Similar effect was obtained for itraconazole and caspofungin while fluconazole was inactive.

Table 7

	Aspergillus fumigatus		*Aspergillus flavus*	
	MEC (μg/ml)	ZOI (mm)	MEC (μg/ml)	ZOI (mm)
Methanol	0	0	0	0
Compound 2(a)	2.5	2.7	2.5	2.7
Itraconazole	1.0	1.7	0.5	1.7
Caspofungin	2.5	0.7	2.5	0.7
Fluconazole	0	0	0	0

MEC : mimimum effective concentration

ZOI : zone of inhibition of fungal growth calculated for each MEC

Example 6. Evaluation of Antifungal Activity of Compound 2(a) in a Mouse Model of Disseminated Candidiasis

Compound 2(a) was provided as a dry powder with an estimated purity of 95+%. Fungizone (amphotericin B desoxycholate, to be used as a comparitor), was also provided as a dry powder with an estimated purity of 95+%. The compound 2(a) and Fungizone were stored as dry powders at -80°C until the day of administration.

Female mice (species *Mus musculus*, strain CD-1, Charles River) with body weight range of 22-24 g were used in the study. The animals were observed for 3 days before treatment. All animal experiments were performed at the Ste-Justine Hospital (Montreal, Quebec) according to ethical guidelines of animal experimentation of the ethical committee of the hospital. During the study, dead or apparently sick animals were promptly removed and sick mice were euthanized upon removal from the cage.

The animals were maintained in rooms under controlled conditions of temperature (23±2°C), humidity (45±5%), photoperiodicity (12 hrs light / 12 hrs dark) and air exchange. The animals were housed in polycarbonate cages (4/single cage) equipped to provide food and water. Sterile wood

shavings were used for animal bedding and the bedding was replaced every other day. Food (Harlam Tecklab, Canada) and autoclaved tap water was provided *ab libitum*, the food being placed in the metal lid on top of the cage. Water bottles were equipped with rubber stoppers and sipper tubes and were cleaned, sterilized and replaced once a week.

Six groups of mice (10 mice per group) were infected intravenously with 3 x 10^6 CFU of *C. albicans* SC5314 as previously described (see Dubois, N., et al., *Microbiology* 1998, 144: 2299-2310). Twenty-four hours after infection, each individual group of mice was treated with Compound 2(a) (1 or 3 mg/kg i.p.), Fungizone (0.25, 0.5 or 1 mg/kg i.p.) as comparitor, or sham-treated with sterile water containing 5% dextrose and 3% DMSO. Each animal received 100 µl of test solution.

The treatment regimen was repeated once daily for a total of 4 days. The mice were observed twice daily for signs of morbidity over 21 days. Moribund animals were scored as non-survivors and euthanized by CO_2 inhalation. The Kaplan and Meier product limit estimate was used to analyze survival data and plot the survival function.

Table 8: Survival Rates Over Time After Inoculation with Compound 2(a) and Fungizone

Groups	Treatment	Dose (mg/kg)	Median survival
1	Vehicle	-	5 days
2	Compound 2(a)	1.0	8.5 days
3	Compound 2(a)	3.0	20 days
4	Fungizone	0.25	>21 days
5	Fungizone	0.5	>21 days
6	Fungizone	1.0	>21 days

As indicated in Table 8, compound 2(a) has *in vivo* antifungal activity similar to a dose of 0.25 mg/kg of Fungizone and increases 4-fold the median survival time of infected mice.

The data (percent survival versus days post-inoculation) was plotted; the resulting graph is shown in Figure 16.

Example 7. In Vitro Antitumor activity of Compound 2(a)

In vitro antipoliferative study of Compound 2a was performed by the National Cancer Institute (National Institutes of Health, Bethesda, Maryland, USA) against a panel of cancer cell lines in order to determine the concentrations needed to obtain a 50% inhibition of cell proliferation (IC_{50}). The operation of this unique screen utilizes 60 different human tumor cell lines, representing leukemia, melanoma, and cancers of the lung, colon, brain, ovary, breast prostate and kidney. Compound 2(a) was provided as a lyophilized powder with an estimated purity of 90+%. The compound was stored at -20°C until day of use.

The human tumor cell lines of the cancer-screening panel were grown in RPMI 1640 medium containing 5% fetal bovine serum and 2 mM L-glutamine. For a typical screening experiment, cells were inoculated into 96 well microtiter plates in 100 µl at plating densities ranging from 5000 to 40,000 cells/well depending on the doubling time of individual cell lines (Table 8). After cell inoculation, the microtiter plates were incubated at 37 °C, under 5% CO_2, 95% air and 100% relative humidity for 24 hours prior to addition of the experimental drugs.

After 24 hours, two plates of each cell line were fixed *in situ* with TCA, to represent a measurement of the cell population for each cell line at the time of drug addition (Tz). Compound 2(a) was solubilized in dimethyl sulfoxide at 400-fold the desired final maximum test concentration and stored frozen prior to use. At the time of drug addition, an aliquot of frozen concentrate was thawed and diluted to twice the desired final maximum test concentration with complete medium containing 50 µg/ml gentamicin. Additional four, serial dilutions were made to provide a total of five drug concentrations plus control. Aliquots of 100 µl of these different drug dilutions were added to the appropriate microtiter wells already containing 100 µl of medium, resulting in the required final drug concentrations (2.5×10^{-5} M to 2.5×10^{-9} M).

Following drug addition, the plates were incubated for an additional 48 hours at 37°C, 5 % CO_2, 95 % air, and 100 % relative humidity. For adherent

cells, the assay was terminated by the addition of cold TCA. Cells were fixed *in situ* by the gentle addition of 50 µl of cold 50 % (w/v) TCA (final concentration, 10 % TCA) and incubation for 60 minutes at 4°C. The supernatant was discarded, and the plates were washed five times with tap water and air-dried. Sulforhodamine B (SRB) solution (100 µl) at 0.4 % (w/v) in 1 % acetic acid was added to each well, and plates were incubated for 10 minutes at room temperature. After staining, unbound dye was removed by washing five times with 1 % acetic acid and the plates were air-dried. Bound stain was subsequently solubilized with 10 mM trizma base, and the absorbance was read on an automated plate reader at a wavelength of 515 nm. For suspension cells, the methodology was the same except that the assay was terminated by fixing settled cells at the bottom of the wells by gently adding 50 µl of 80 % TCA (final concentration, 16 % TCA).

The growth inhibitory power of compound 2(a) was measured by NCI utilizing the GI_{50} value, rather than the classical IC_{50} value. The GI_{50} value emphasizes the correction for the cell count at time zero and, using the seven adsorbance measurements [time zero (Tz), control growth (C), and the test growth in the presence of drug at each of the five concentration levels (Ti)], GI_{50} is calculated as [(Ti - Tz) / (C - Tz) x 100 = -50. which is the drug concentration resulting in a 50% reduction in the net protein increase (as measured by SRB staining) in control cells during the drug incubation. The GI_{50} values for compound 2(a) for the various cell lines tested are presented in Table 9 below.

Table 9: NCI Developmental Therapeutics Program In-Vitro Testing Results for Compound 2(a)

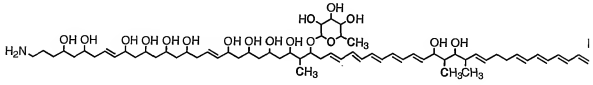
Cell Line	Panel name	Inoculation density (no. of cells per well)	GI_{50} (x 10^{-6}, unless otherwise indicated)
K-562	Leukemia	5000	9.18
MOLT-4	Leukemia	30,000	5.57
A549/ATCC	Non-small cell lung cancer	7500	4.09

EKVX	Non-small cell lung cancer	20,000	5.87
HOP-62	Non-small cell lung cancer	10,000	6.83
HOP-92	Non-small cell lung cancer	20,000	9.77×10^{-8}
NCI-H226	Non-small cell lung cancer	20,000	3.10
NCI-H23	Non-small cell lung cancer	20,000	4.25
NCI-H322M	Non-small cell lung cancer	20,000	3.48
NCI-H460	Non-small cell lung cancer	7500	3.83
NCI-H522	Non-small cell lung cancer	20,000	2.80
COLO 205	Colon cancer	15,000	5.00
HCC-2998	Colon cancer	15,000	6.03×10^{-8}
HCT-116	Colon cancer	5000	4.18
HCT-15	Colon cancer	10,000	3.25
HT29	Colon cancer	5000	6.36
KM12	Colon cancer	15,000	2.76
SW-620	Colon cancer	10,000	5.35
SF-268	CNS cancer	15,000	3.64
SF-295	CNS cancer	10,000	3.91
SNB-19	CNS cancer	15,000	5.58
SNB-75	CNS cancer	20,000	3.87
U251	CNS cancer	7500	3.65
LOX IMVI	Melanoma	7500	3.73
MALME-3M	Melanoma	20,000	2.40
M14	Melanoma	15,000	4.15
SK-MEL-2	Melanoma	20,000	4.34
SK-MEL-28	Melanoma	10,000	6.75
SK-MEL-5	Melanoma	10,000	4.16
UACC-257	Melanoma	20,000	3.74
UACC-62	Melanoma	10,000	2.68
IGROV1	Ovarian cancer	10,000	2.95
OVCAR-3	Ovarian cancer	10,000	3.40
OVCAR-4	Ovarian cancer	15,000	4.48
OVCAR-5	Ovarian cancer	20,000	4.00
OVCAR-8	Ovarian cancer	10,000	4.34
SK-OV-3	Ovarian cancer	20,000	7.94
786-0	Renal cancer	10,000	3.07

A498	Renal cancer	25,000	4.82
ACHN	Renal cancer	10,000	2.96
CAKI-1	Renal cancer	10,000	2.99
RXF 393	Renal cancer	15,000	1.20
SN12C	Renal cancer	15,000	1.38×10^{-7}
TK-10	Renal cancer	15,000	3.32
UO-31	Renal cancer	15,000	3.65
PC-3	Prostate cancer	7500	2.66
DU-145	Prostate cancer	10,000	3.78
MCF7	Breast cancer	10,000	4.22
NCI/ADR-RES	Breast cancer	15,000	4.76
MDA-MB-	Breast cancer	20,000	3.38
MDA-MB-435	Breast cancer	15,000	3.26
BT-549	Breast cancer	20,000	4.59
T-47D	Breast cancer	20,000	6.00

The results indicate that compound 2(a) is effective against all the human tumor cell lines that have been assayed in the NCI screening panel suggesting a broad anticancer activity against several types of human cancer. In fact, the GI50 calculated for all cell lines was lower than 10 x10-6 M, a significant level of pharmacological activity for anticancer drugs, and in some cases reached the nanomolar or picomolar level (SN12C/renal carcinoma; HOP92/non-small cell lung carcinoma; HCC2998/colon carcinoma).

Example 8 Activation of inactive domains in the polyketide synthase system



The gene cluster encoding the Compound 2(a) derived from *Streptomyces aizunensis* strain NRRL B-11277 is genetically modified to reactivate the ketoreductase (KR) domain, which is encoded in the ORF 13 module 12. This modification results in the conversion of the central carbonyl group adjacent to the sugar molecule of Compound 2(a), to a hydroxyl group (as shown in Figure 12a).

In the compound 2(a) locus, the KR domain present in ORF 13, module 12 is inactive. To provide for the compound of Example 7 the KR domain is reactivated or swapped for an active KR domain. Reactivation of the KR domain requires diagnosis of the integrity of critical active site residues necessary for a functional KR domain. The active site residues can be divided into those required for co-enzyme activation of the KR enzyme and those for catalysis. Experiments identifying the specific residues for ketoreductase activity [Ried *et. al.* Biochemistry 2003, 42:72-79; Udo *et.al.*, Biochemistry, 1997, 36:34-40] reveal that functional KR coenzyme binding site residues include glycine (G), glycine (G), glycine (G), alanine (A) and the functional KR active site residues include serine (S), tyrosine (Y) and asparagine (N). These residues are highlighted in Figures 6a and 6b. The sequence of the KR domain in the compound 2(a) locus shows that the coenzyme active site residues are glycine (G), glycine (G), glycine (G), alanine (A) indicating that this site is indeed active. However, the amino acid residues found in the KR site responsible for catalytic activity are serine (S), glutamine (Q) and asparagine (N) indicating that the catalytic site is likely to be inactive. This observation is confirmed by the fact compound 2(a) contains a carbonyl group at that specific position (Figure 10, module 12). Modification of the codon encoding glutamine to a codon encoding tyrosine provides for an active site residue required for functional ketoreduction of PKS monomers. This results in an altered nucleic acid sequence of the compound 2(a) locus used to modify a suitable host cell to produce the compound 2(a) variant of Example 7 as shown in Figure 12a.

The modification of glutamine to tyrosine may be introduced using a mismatched primer that hybridizes to the native nucleotide sequence at a temperature below the melting temperature of the mismatched duplex. The primer is kept specific by keeping primer length and base composition within narrow limits and keeping the mutant base centrally located as described in Zoller and Smith' Methods in Enzymol. (1983) 100:468. Primer extension is achieved using DNA polymerase. The product is cloned and positive clones containing the mutated DNA, derived by segregation of the primer extended strand, are selected. Selection is made using the mutant primer as a

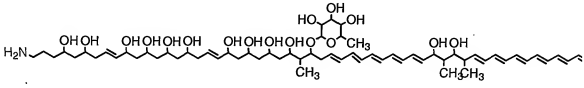
hybridization probe (Dalbie-McFarland et al Proc. Natl. Acad Sci. USA (1982) 79:6409).

Another method to generate the compound of Example 7 involves swapping the inactive ketoreductase domain from the gene locus of the compound 2(a) (ORF 13 module 12) with an active ketoreductase domain from the same or different locus. Example of domains within the same locus suitable for swapping include the active ketoreductases that occur in the modules that encode the incorporation of methyl malonate extender units, namely ORF 16 modules 19 or 20. Swapping of acyltransferase domains between PKS loci has been demonstrated by Oliynyk *et.al.* Chem Biol, 1996, 3(10):833-9, wherein the gene encoding the acyltransferase domain in 6-deoxyerythronolide (DEBS) module 1 is swapped with the gene encoding the rapamycin module 2 acyltransferase resulting in the synthesis of novel triketides since the two acyltransferases had different acyl specificities. In Hans *et.al.* J Am Chem Soc, 2003, 125(18):5366-74, the kinetic aspects of product formation as a consequence of acyltransferase domain swaps is taught.

Swapping of domains is achieved using techniques developed by Kao *et.al.* Science, 1994, 265:509-512. The genetic strategy utilizes derivatives of pMAK705 to permit in vivo recombination between a temperature sensitive donor plasmid and a recipient shuttle vector by means of a double recombination event in *E.coli*. An Amp^R Tc^R recipient subclone of the regions flanking the domain to be swapped is made, pCK5, containing 1kb of flanking sequence from either flank. Endonuclease restriction sites are introduced at the boundaries of the domain, *PstI* at 3' end of the left flank and *XbaI* at the 5' end of the right flank. Subclones pCK6 Cm^R of the domains to be swapped are generated and endonuclease restriction sites are introduced into the boundaries of the domain. The restriction site *PstI* is introduced at the 5' boundary of the KR domain and an *XbaI* site at the 3' boundary of the domain. Restriction sites are introduced into subclones by PCR mutagenesis. The fragment containing the domain is excised and ligated into the temperature sensitive Cm^R donor plasmid, pCK6. The recipient plasmid is generated by in vivo recombination of the plasmid in the host strain using the selection method

outlined by Kao *et.al., supra*. After selection recombinant strains are produced with the domain of interest replacing the original domain.

Example 9 Inactivation of functional domains within the polyketide synthase system



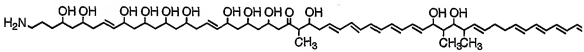
The gene locus encoding Compound 2(a) derived from a *Streptomyces aizunensis* strain is genetically modified to inactivate the enoyl reductase (ER) domain in the ORF 17 module 22. Inactivation of this domain abolishes the conversion the double bond to the single bond between the acyl units incorporated by modules 21 and 22 of Compound 2(a) (as shown in Figure 12e).

Generating the compound of Example 8 is achieved through insertional inactivation by double crossover techniques developed by Oh and Chater, 1997, Journal of Bacteriology 179:122-127. Examples of insertional inactivation of genes involved in polyketide biosynthesis in *Streptomyces* are well known in the art. Arrowsmith *et.al.*, 1992, Mol Gen Genet 234:254-264, used these techniques to identify the role of a cassette of secondary metabolic genes in the production of monensin by *Streptomyces cinnamonensis*. Paradkar, *et.al.*, 2001, Appl Environ Microbiol 67:2292-7, inactivated the *lat* gene encoding for lysine aminotransferase to disrupt the first step in the cephamycin pathway to block production of cephamycin C in *Streptomyces clavuligerus*. Similarly, these authors inactivated the *cvm*1 gene involved in late stage antipodal clavam synthesis.

Methods used to inactivate domains in polyketide systems include domain swapping as described in Example 7 as well as targeted disruption by insertional gene inactivation. For this, a replicative plasmid-mediated homologous recombination is applied to *Streptomyces aizunensis*. Plasmids for homologous recombination are constructed by cloning a kanamycin resistance marker between the left and right flanking regions of the genes to be modified. Such a construct is cloned into a delivery plasmid that is marked with thiostrepton resistance producing a disruption plasmid. This plasmid is

introduced into *Streptomyces aizunensis* by either PEG-mediated protoplast transformation, by electroporation or by natural infection with a phage (Keiser *et al* (2000) Practical Streptomyces genetics, John Innes Foundation, Norwich). The spores from individual transformants or transconjugants are cultured on non-selective plates to induce recombination. The cycle is repeated three times to enhance the opportunity for recombination. Crossovers yielding targeted gene recombinants are then selected and screened using kanamycin and thiostrepton for single crossovers and kanamycin for double crossovers. Replica plating and southern hybridization are used to confirm the double crossover inactivation (Keiser *et al* (2000) *supra*.).

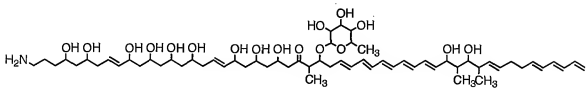
Example 10 Inactivation of the glycosyltransferase activity



Inactivation of the glycosyltransferase gene (GTFA) encoding ORF 9 of the compound 2(a) locus (as shown in Figure 12b) provides for the compound of this example. The inactivation of the GFTA disrupts the transfer of the sugar moiety onto the backbone of Compound 2(a). The absence of the sugar moiety results in a non-glycosylated form of Compound 2(a). Insertional inactivation of GTFA genes in polyketide biosynthesis in *Streptomyces* is known in the art. Blanco *et.al.*, 2000, Mol Gen Genet 262:991-1000, identified two genes of the mithramycin biosynthetic gene cluster as glycosyltransferases by the production of a non-glycosylated mithramycin upon inactivation of these genes. A similar observation was made by Chen *et.al.*, Gene, 2001, 263:255-64 investigating genes responsible for glycosylation in the biosynthetic pathways encoding pikromycin, narbomycin, methymycin and neomethymycin.

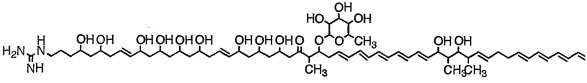
Targeted inactivation of the glycosyltransferase activity is achieved using the method of insertional gene disruption as described in Example 8.

Example 11 **Elimination of the aminohydroxycyclopentenone unit**



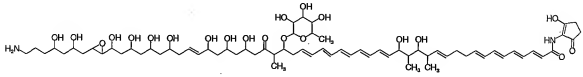
Elimination of the terminal aminohydroxycyclopentenone unit may be accomplished by inactivation of any one of the following three ORFs of the compound 2(a) locus. First, disruption of ORF 35 results in the inactivation of the acyltransferase (AYTP) activity (as shown in Figure 12c) that abolishes condensation of succinyl-CoA and glycine to form 5-aminolevulinate. Second, disruption of ORF 36 results in the inactivation of acyl CoA ligase (CALB) preventing the conversion of 5-aminolevulinate to 5-aminolevulinate-CoA which cyclizes to form aminohydroxycyclopentenone. Third, disruption of ORF 34 (ADSN) prevents transfer of the aminohydroxycyclopentenone unit to the polyketide chain. Thus, the compound of Example 10 is provided by genetically modifying at least one of ORFs 34, 35 and 36. Methods used for insertional inactivation of all three genes are described in Example 9.

Example 12 **Replacement of the terminal amine group with a guanidino group**



The replacement of the terminal amine with a guanidino group may be accomplished by the insertional inactivation of ORF 33 (ADHY) using the methods described in Example 9. The inactivation of ORF 33 ADHY (as shown in Figure 12d) disrupts the synthesis of gamma-amino butanamide leading to the accumulation of 4-guanidino butanamide. The accumulated 4-guanidino butanamide is converted by ORF 27 CALB to 4-guanidino butyryl-CoA which is then attached onto the polyketide synthase enzyme (ORF 10, module 0 as shown in Figure 10b) through the action of ORF 19 (AYTF).

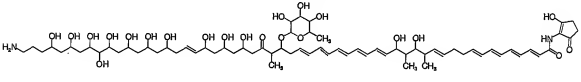
Example 13: Synthesis of Compound 2(b) by epoxidation of Compound 2(a)



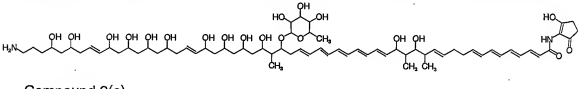
Compound 2(b)

To a mixture of Compound 2(a) dissolved in tetrahydrofuran (THF) is added 1 equivalent of *meta*-chloroperbenzoic acid. The reaction is cooled in an ice bath and stirred at 0 °C for 1-2 hours. The reaction mixture is then evaporated to dryness, re-dissolved in methanol and subjected to liquid chromatography on a column of Sephadex LH-20 to isolate the Compound 2(b).

The epoxide group of Compound 2(b) may be hydrolyzed by treatment of Compound 2(b) with small quantity of aqueous hydrochloric acid (1.0 N), thereby forming the corresponding diol of the formula:



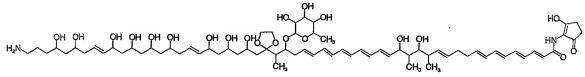
Example 14: Synthesis of Compound 2(c) by Reduction of 31-oxo group



Compound 2(c)

A solution of Compound 2(a) in acetonitrile is treated with 1.5 equivalents of $NaCNBH_3$. The reaction is stirred at room temperature for 1 hour. The reaction mixture is then concentrated to dryness and then taken up into methanol. The mixture is filtered and the filtrate is subjected to liquid chromatography on a column of Sephadex LH-20 to isolate the Compound 2(c). Alternatively, the reduction of the oxo group at the 31-position may be done using lithium borohydride ($LiBH_4$).

Example 15: Synthesis of Compound 2(d) by addition of acetal ring at the 31-position

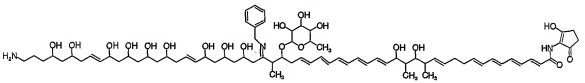


Compound 2(d)

A solution of Compound 2(a) in tetrahydrofuran is treated with 3 equivalents of 2,2-dimethyl-1,3-dioxacyclopentane in the presence of a trace amount of toluene sulfonic acid. The reaction is stirred overnight at room temperature, evaporated to dryness and taken up into dry THF, followed by purification by liquid chromatography on a column of Sephadex LH-20. The 2,2-dimethyl-1,3-dioxacyclopentane may be synthesized by reaction of acetone with ethylene glycol in the presence of a trace of toluene sulfonic acid, over molecular sieves to remove water.

Alternatively, the addition of an acetal ring at the 31-position may be accomplished by reaction of Compound 2(a) with an excess of ethylene glycol in the presence of a trace of toluene sulfonic acid. The reaction may be conducted over molecular sieves to remove water.

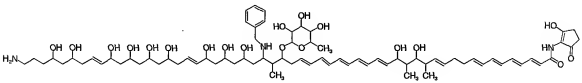
Example 16: Synthesis of Compound 2(e)



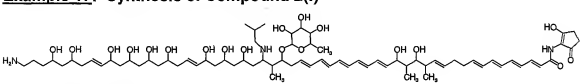
Compound 2(e)

To a solution of Compound 2(a) in benzene or toluene is added 10 equivalents of benzylamine. The reaction is stirred at room temperature overnight. The reaction may be conducted over molecular sieves to remove water; alternatively, the water may be removed under reflux as an azeotrope with benzene or toluene using a Dean-Stark trap. The reaction mixture is concentrated under vacuum and residual reagent is removed by high vacuum at room temperature overnight.

The carbon-nitrogen double bond of Compound 2(e) may be reduced to the amine by reaction of Compound 2(e) with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to form a compound of the structure:



Example 17: Synthesis of Compound 2(f)

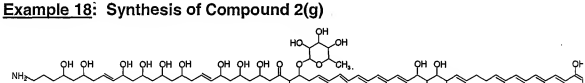


Compound 2(f)

To a solution of one equivalent of Compound 2(a) in acetonitrile is added ten equivalents of isobutylamine. The reaction is stirred at room temperature for two hours. Benzene (1/10 volume) is added and the mixture is concentrated to dryness under vacuum on a rotary evaporator.

The Schiff base is then treated with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to reduce the carbon-nitrogen double bond of the imine to the amine, to form the compound 2(f).

Example 18: Synthesis of Compound 2(g)

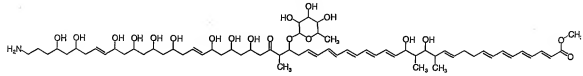


Compound 2(g)

Compound 2(g) may be synthesized biosynthetically as described in Example 9. Alternatively, Compound 2(g) may be prepared by hydrolysis of Compound 2(a). This is accomplished by treatment of Compound 2(a) in diethylether/THF with Meerwein's reagent (triethyloxonium tetrafluoroborate) for two hours at room temperature followed by cooling to –20 °C and dropwise addition of aqueous acetic acid in THF. The reaction mixture is stirred for 20 minutes during which time it is allowed to come to room temperature. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is

added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(g).

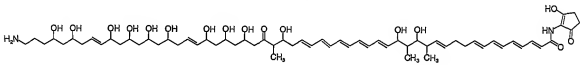
Example 19: Synthesis of Compound 2(h)



Compound 2(h)

To a solution of 0.1 equivalents of Compound 2(g) in methanol is added 0.5 equivalents of diazomethane in diethyl ether. The reaction mixture is allowed to stand at room temperature overnight, and then the solvent is removed under vacuum to give compound 2(h).

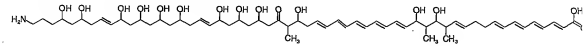
Example 20: Synthesis of Compound 2(i)



Compound 2(i)

A solution of Compound 2(a) in methanol is treated with an equal volume of 0.1N HCl, and the reaction mixture is stirred overnight at room temperature. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(i).

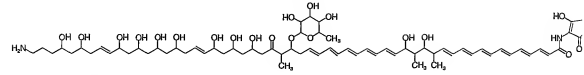
Example 21: Synthesis of Compound 2(j)



Compound 2(j)

Compound 2(j) is prepared by hydrolysis of compound 2(g). The hydrolysis may carried out in the same way that compound 2(a) is hydrolysed to compound 2(i) as described in Example 19 above.

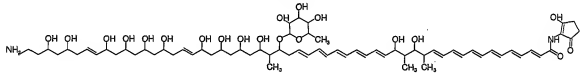
Example 22: Synthesis of Compound 2(k)



Compound 2(k)

Compound 2(k) is prepared biosynthetically by inactivation of the enoyl reductase as described in Example 8.

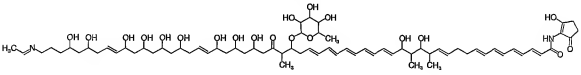
Example 23: Synthesis of Compound 2(l)



Compound 2(l)

A solution of Compound 2(k) in acetonitrile is treated with 1.5 equivalents of $NaCNBH_3$. The reaction is stirred at room temperature for 1 hour. The reaction mixture is then concentrated to dryness and then taken up into methanol. The mixture is filtered and the filtrate is subjected to liquid chromatography on a column of Sephadex LH-20 to isolate the Compound 2(l). Alternatively, the reduction of the oxo group at the 31-position may be done using lithium borohydride ($LiBH_4$).

Example 24: Synthesis of Compound 2(m)

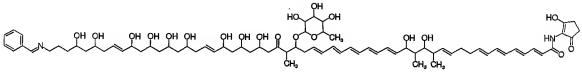


Compound 2(m)

A solution of 10 equivalents of Compound 2(a) in acetonitrile is treated with one equivalent of acetaldehyde. The reaction is stirred at room temperature for two hours. Benzene (1/10 volume) is added and the mixture is concentrated to dryness under vacuum on a rotary evaporator to give the compound 2(m).

Compound 2(m) may be treated with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to reduce the carbon-nitrogen double bond of the imine to the amine.

Example 25: Synthesis of Compound 2(n)

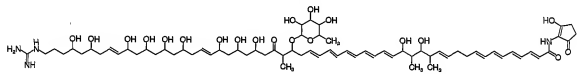


Compound 2(n)

A solution of 10 equivalents of Compound 2(a) in acetonitrile is treated with one equivalent of benzaldehyde. The reaction is stirred at room temperature for two hours. Benzene (1/10 volume) is added and the mixture is concentrated to dryness under vacuum on a rotary evaporator to give the compound 2(n).

Compound 2(n) may be treated with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to reduce the carbon-nitrogen double bond of the imine to the amine.

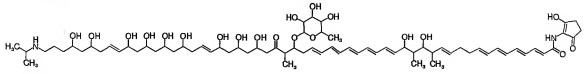
Example 26: Synthesis of Compound 2(o)



Compound 2(o)

A solution of Compound 2(a) in tetrahydrofuran is treated with one equivalent of cyanamide. The reaction mixture is stirred at room temperature overnight. Solvent is removed from the reaction mixture under vacuum to give compound 2(o).

Example 27: Synthesis of Compound 2(p)

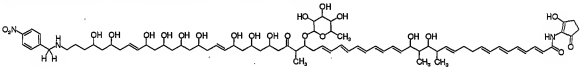


Compound 2(p)

To a solution of 10 equivalents of Compound 2(a) in acetonitrile is added 1 equivalent of acetone. The reaction is stirred at room temperature for two hours. Benzene (1/10 volume) is added and the mixture is concentrated to dryness under vacuum on a rotary evaporator.

The resulting Schiff base imine is then treated with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to reduce the carbon-nitrogen double bond of the imine to the amine, to form the compound 2(p).

Example 28: Synthesis of Compound 2(q)

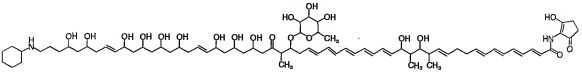


Compound 2(q)

To a solution of 10 equivalents of Compound 2(a) in acetonitrile is added 1 equivalent of 4-nitrobenzaldehyde. The reaction is stirred at room temperature for two hours. Benzene (1/10 volume) is added and the mixture is concentrated to dryness under vacuum on a rotary evaporator.

The resulting Schiff base imine is then treated with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to reduce the carbon-nitrogen double bond of the imine to the amine, to form the compound 2(q).

Example 29: Synthesis of Compound 2(r)

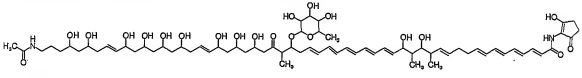


Compound 2(r)

To a solution of 10 equivalents of Compound 2(a) in acetonitrile is added 1 equivalent of cyclohexylformaldehyde. The reaction is stirred at room temperature for two hours. Benzene (1/10 volume) is added and the mixture is concentrated to dryness under vacuum on a rotary evaporator.

The resulting Schiff base imine is then treated with $NaCNBH_3$ or $LiBH_4$ (1.5 equivalents) in acetonitrile, to reduce the carbon-nitrogen double bond of the imine to the amine, to form the compound 2(r).

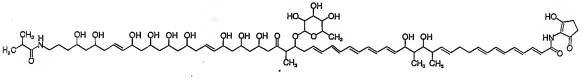
Example 30: Synthesis of Compound 2(s)



Compound 2(s)

To a solution of Compound 2(a) in tetrahydrofuan is added one equivalent of acetic anhydride and two equivalents of triethylamine. The reaction is stirred at room temperature for two hours. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(s).

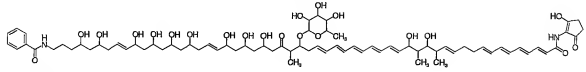
Example 31: Synthesis of Compound 2(t)



Compound 2(t)

To a solution of Compound 2(a) in is added one equivalent of isobutyrl anhydride and two equivalents of triethylamine. The reaction is stirred at room temperature for two hours. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(t).

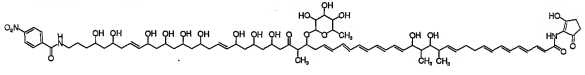
Example 32: Synthesis of Compound 2(u)



Compound 2(u)

To a solution of Compound 2(a) in is added one equivalent of benzoic anhydride and two equivalents of triethylamine. The reaction is stirred at room temperature for two hours. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(u).

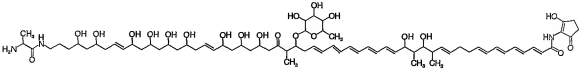
Example 33: Synthesis of Compound 2(v)



Compound 2(v)

To a solution of Compound 2(a) in is added one equivalent of p-nitrobenzoic anhydride and two equivalents of triethylamine. The reaction is stirred at room temperature for two hours. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(v).

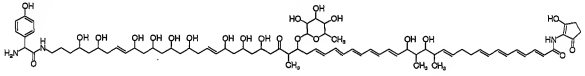
Example 34: Synthesis of Compound 2(w)



Compound 2(w)

A solution of Compound 2(a) is reacted with 1 equivalent of N-protected alanine active ester. The amino group of alanine is protected by reacting alanine with DCC (dicyclohexyldicarbodiimide) or EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and the carboxylic acid group is converted to an active ester such as an N-hydroxysuccinimide ester. The N-protected active ester is added to Compound 2(a) in an inert solvent such as tetrahydrofuran. The mixture is warmed under reflux for one hour. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(w).

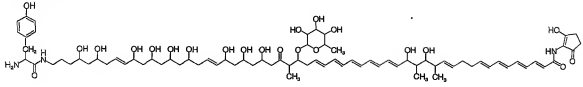
Example 35: Synthesis of Compound 2(x)



Compound 2(x)

A solution of Compound 2(a) is reacted with 1 equivalent of N-protected *para*-hydroxyphenyl glycine active ester. The amino group of the *para*-hydroxyphenyl glycine is protected by reacting alanine with DCC (dicyclohexyldicarbodiimide) or EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and the carboxylic acid group is converted to an active ester such as an N-hydroxysuccinimide ester. The N-protected active ester is added to Compound 2(a) in an inert solvent such as tetrahydrofuran. The mixture is warmed under reflux for one hour. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(x).

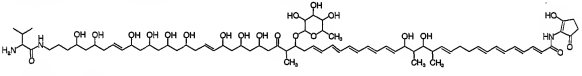
Example 36: Synthesis of Compound 2(y)



Compound 2(y)

A solution of Compound 2(a) is reacted with 1 equivalent of N-protected tyrosine active ester. The amino group of tyrosine is protected by reacting alanine with DCC (dicyclohexyldicarbodiimide) or EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and the carboxylic acid group is converted to an active ester such as an N-hydroxysuccinimide ester. The N-protected active ester is added to Compound 2(a) in an inert solvent such as tetrahydrofuran. The mixture is warmed under reflux for one hour. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(y).

Example 37: Synthesis of Compound 2(z)

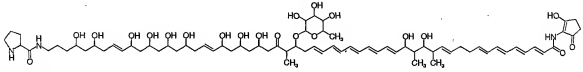


Compound 2(z)

A solution of Compound 2(a) is reacted with 1 equivalent of N-protected valine active ester. The amino group of valine is protected by reacting alanine with DCC (dicyclohexyldicarbodiimide) or EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and the carboxylic acid group is converted to an active ester such as an N-hydroxysuccinimide ester. The N-protected active ester is added to Compound 2(a) in an inert solvent such as tetrahydrofuran. The mixture is warmed under reflux for one hour. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well

with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(z).

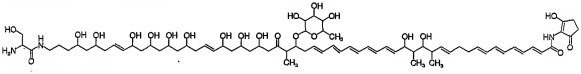
Example 38: Synthesis of Compound 2(aa)



Compound 2(aa)

A solution of Compound 2(a) is reacted with 1 equivalent of N-protected proline active ester. The amino group of proline is protected by reacting alanine with DCC (dicyclohexyldicarbodiimide) or EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and the carboxylic acid group is converted to an active ester such as an N-hydroxysuccinimide ester. The N-protected active ester is added to Compound 2(a) in an inert solvent such as tetrahydrofuran. The mixture is warmed under reflux for one hour. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(aa).

Example 39: Synthesis of Compound 2(ab)



Compound 2(ab)

A solution of Compound 2(a) is reacted with 1 equivalent of N-protected serine active ester. The amino group of serine is protected by reacting alanine with DCC (dicyclohexyldicarbodiimide) or EDC (1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide) and the carboxylic acid group is converted to an active ester such as an N-hydroxysuccinimide ester. The N-protected active ester is added to Compound 2(a) in an inert solvent such as tetrahydrofuran. The mixture is warmed under reflux for one hour. The mixture is then diluted with water (2 volumes) and HP-20 polystyrene resin is

added. The mixture is stirred for 30 minutes, filtered, the resin is washed well with water, and the product is eluted with 100% ethanol. The elutes are concentrated under vacuum to give compound 2(ab).

Example 40: Compound 2(a) for the treatment of cardiovascular disorders

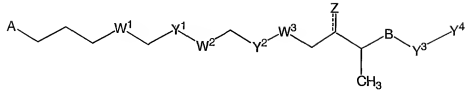
Polyene compounds are not generally absorbed from the gastrointestinal tract and exhibit hypocholesterolemic properties by binding cholesterol in the gastrointestinal tract following oral administration. The hypocholesterolemic properties of polyene compounds was first demonstrated by studies in dogs (Schaffner,C.P. and Gordon H.W. The hypocholesterolemic activity of orally administered polyene macrolides. P.N.A.S. 61:36-41, 1968.). In another study with chickens, small amounts of polyene compounds in the diet led to the inhibition of enterohepatic cholesterol circulation, increased fecal lipid excretion and reduced atherogenesis (Fisher, H., Griminger P. and Siller W. Effect of candicidin on plasma cholesterol and avian atherosclerosis. Proceedings of the Society for Experimental Biology and Medicine, 145: 836-839, 1974). The beneficial effects of orally administered polyene compounds on cholesterol-lipid metabolism is not species-dependent as it was demonstrated in several species including humans, rats, dogs and chickens (Pagliano FM, Correction of hyperdyslipidemia using polyene-structure substances. Controlled clinical trial. Arch Sci Med (Torino). 136: 303-308, 1979; Barbaro A. and Casella G. Action of a polyene macrolide on hyperdislipidaemic disorders. Archivio per Scienze Mediche 137: 211-216, 1980; Singhal, A.K., Mosbach, E.H. and Schaffner, C.P. Effect of candicidin on cholesterol and bile acid metabolism in the rat. Lipids, 16: 423-426, 1981.).

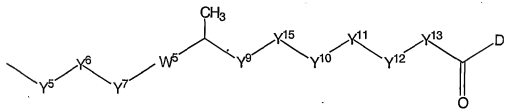
The therapeutic potential of compound 2(a) for the treatment of cardiovascular disorders such as high cholesterol, dyslipidemia and atherosclerosis is demonstrated by measuring the effects of oral administration of compound 2(a) to rabbits. New Zealand rabbits are maintained under controlled light and temperature conditions and fed for several weeks with two different diets: normal rabbit chow (control) and a diet

containing 0.5 to 1% cholesterol to induce hypercholesterolemia. Rabbits are administered compound 2(a) (3, 10, 30 mg/kg) or vehicle by oral gavage daily for up to one month. Food intake and rabbit weight is measured daily for the duration of the experiment. Blood samples to measure cholesterol, lipoproteins and triglycerides are collected through a catheter inserted in the ear artery in the beginning and at the end of the experiment as well as every 4 days for the duration of the experiment. Serum cholesterol, lipoproteins and triglycerides are measured by enzymatic assays employing commercial kits as specified by the manufacturer (Sigma Chemical Co) and as described in Staprans I, Pan X-M, Rapp JH, Feingold KR. Oxidized cholesterol in the diet accelerates the development of aortic atherosclerosis in cholesterol-fed rabbits. Arteriosclerosis, Thrombosis and Vascular Biology, 18: 977-983, 1998. At the end of the experiment, after collecting the final blood sample, animals are anesthetized and the descending aorta is exposed, excised and processed for histological examination following fixation in formalin. Briefly, paraffin longitudinal or cross sections (five micron) are stained with Sudan black (dying lipids) and counterstained with Masson trichrome. Morphometric quantitative determination of the area of the intima, media and adventitia layers is performed by image analysis. Lipid deposition in the aorta is determined by evaluation of the percentage of the aorta covered by lesions visualized by fat staining. Arterial concentration of cholesterol is measured after extraction of lipids as described in Thiery J, Nebendahl K, Rapp K, Kluge R, Teupser D and Seidel D. Low atherosclerotic response of a strain of rabbits to diet-induced hypercholesterolemia. Arteriosclerosis, Thrombosis and Vascular Biology, 15: 1181-1188, 1995.

What is Claimed is:

1. A compound of Formula I,



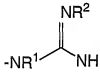


Formula I

or a pharmaceutically acceptable salt thereof;

wherein,

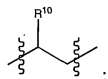
A is selected from the group consisting of $-NR^1R^2$, $-N=CR^1R^2$,



, and


;

R^1, R^2, R^3 and R^4 are each independently selected from the group consisting of H, C_{1-6} alkyl, C_{2-6} alkenyl, C_{3-6} cycloalkyl, C_{2-6} heterocycloalkyl, aryl, heteroaryl and amino acid, wherein said alkyl, alkenyl, aryl and heteroaryl groups are optionally substituted with a group selected from halogen, OH, NO_2, NH_2 or aryl, said aryl being optionally further substituted with one or more groups independently selected from halogen, OH, NO_2 or NH_2;

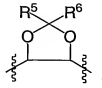
B is selected from: ethene-1,2-diyl or 

;

wherein R^{10} is oxo or OR^{11};

wherein R^{11} is H or a heterocycloalkyl, the heterocycloalkyl being optionally substituted with 1-4 substituents selected from OX, C_{1-3} alkyl and $-O-C(O)R^1$, wherein X is H or, when there are at least two neighboring

substituent groups that are OX, then the X can be a bond such that the two neighboring oxygen groups form a five-membered acetal ring of the formula:



; wherein R^5 and R^6 are each independently selected from the group consisting of H, C_{1-6} alkyl, and C_{2-7} alkenyl;

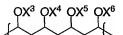
D is selected from: 

, $-NR^{12a}R^{12a}$, and OR^{12}, wherein R^{12} is selected from H, C_{1-6} alkyl optionally substituted with 1 to 2 phenyl groups, wherein the phenyl group is optionally substituted with C_{1-6} alkyl or halo;

R^{12a} and R^{12a} are each indepedently selected from H, C_{1-6} alkyl, C_{2-6} alkenyl, C_{3-6} cycloalkyl, C_{2-6} heterocycloalkyl, aryl, heteroaryl and amino acid, wherein said alkyl, alkenyl, aryl and heteroaryl are optionally substituted with a group selected from halogen, OH, NO_2, NH_2 or aryl, said aryl being optionally further substituted with one or more groups independently selected from halogen, OH, NO_2 or NH_2;

W^1 is 

;

W^2 is : 

;

W^3 is 

;

W^5 is 

;

X^1, X^2, X^3, X^4, X^5, X^6, X^7, X^8, X^9, X^{12} and X^{13} is each independently selected from H, -C(O)-R^7 and a bond such that when any of two neighboring X^1, X^2, X^3, X^4, X^5, X^6, X^7, X^8, X^9, X^{12} and X^{13} is a bond then the two neighboring oxygen atoms and their attached carbon atoms together form a six-membered acetal ring of the formula:

wherein R^5, R^6 and R^7 are each independently selected from H, C_{1-6} alkyl, C_{2-7} alkenyl;

Y^1, Y^2, Y^3, Y^4, Y^5, Y^6, Y^7, Y^9, Y^{10}, Y^{11}, Y^{12}, Y^{13} and Y^{15} are each independently selected from the group consisting of ethene-1,2-diyl, ethane-1,2-diyl and  , wherein said ethene-1,2-diyl and ethane-1,2-diyl groups are optionally substituted with a methyl group;

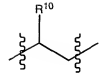
Z is selected from OH, NHR^8, and  and when the dotted line is a bond then Z is oxo, or NR^9;

R^8 is independently selected from H, C_{1-6} alkyl, C_{2-6} alkenyl;

R^9 is C_{1-6} alkyl optionally substituted with aryl.

2. The compound of claim 1, wherein Z is oxo, or a pharmaceutically acceptable salt thereof.

3. The compound of claim 1 or 2, wherein D is  , or a pharmaceutically acceptable salt thereof.

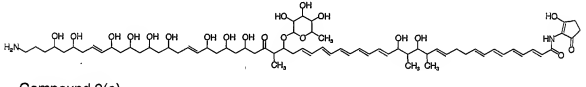
4. The compound of claim 1,2 or 3, wherein B is  or a pharmaceutically acceptable salt thereof.

5. The compound of claim any one of claims 1 to 4, wherein A is -NR^1R^2, or a pharmaceutically acceptable salt thereof.

6. The compound of claim 4, wherein A is  or a pharmaceutically acceptable salt thereof.

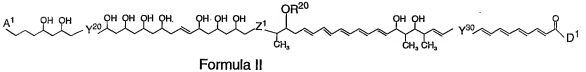
7. The compound of any one of claims 1, 2, 4, 5 or 6, wherein D is  , or a pharmaceutically acceptable salt thereof.

8. A compound of the fomula:



Compound 2(a)

9. A compound of the formula II:



Formula II

wherein A^1 is $-NH_2$, –N=CH-R^{13}, amino acid or -NH-R^{14}, wherein R^{13} is hydrogen or phenyl and R^{14} is selected from the group consisting of isopropyl, 1-(4-nitrophenyl)methyl, cyclohexyl, and wherein said amino acid is attached via its nitrogen atom;


and


wherein R^{15} is selected from the group consisting of methyl, isopropyl, phenyl, 4-nitrophenyl, 1-aminoethyl, 1-amino-1-(4-hydroxyphenyl)methyl, 1-amino-2-(4-hydroxyphenyl)ethyl, 1-amino-2-methylpropyl, 2-pyrrolidinyl and1-amino-2-hydroxyethyl;

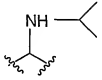
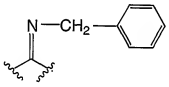
Y^{20} is selected from the group consisting of ethene-1,2-diyl and


;

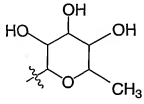
Z^1 is selected from the group consisting of:


,

,

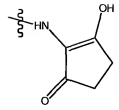
,

and


R^{20} is selected from the group consisting of hydrogen and



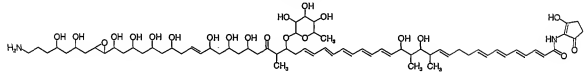
Y^{30} is ethene-1,2-diyl or ethane-1,2-diyl; and

D^1 is hydroxy, methoxy or

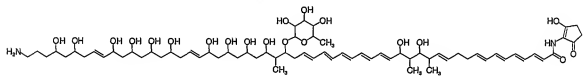


and pharmaceutically acceptable salts thereof.

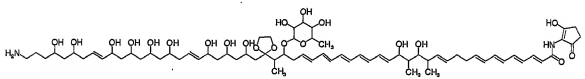
10. A compound selected from the group consisting of:



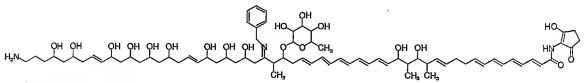
Compound 2(b)



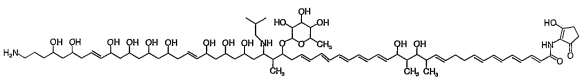
Compound 2(c)



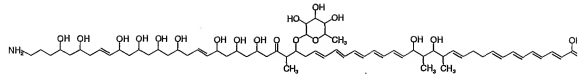
Compound 2(d)



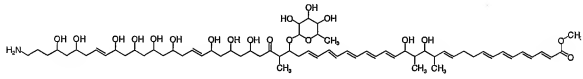
Compound 2(e)



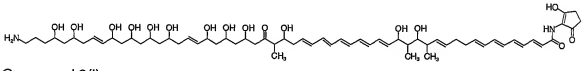
Compound 2(f)



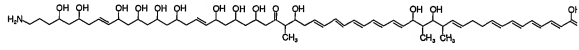
Compound 2(g)



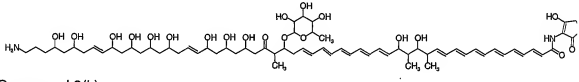
Compound 2(h)



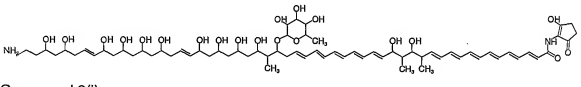
Compound 2(i)



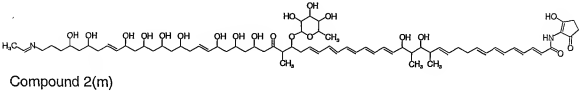
Compound 2(j)



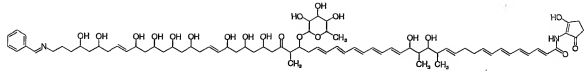
Compound 2(k)



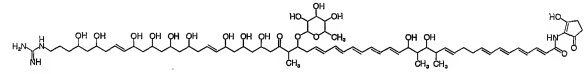
Compound 2(l)



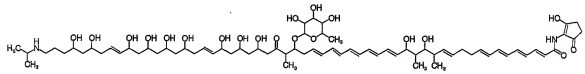
Compound 2(m)



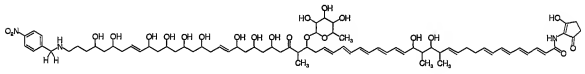
Compound 2(n)



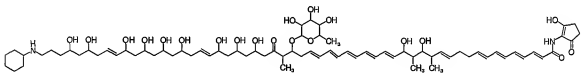
Compound 2(o)



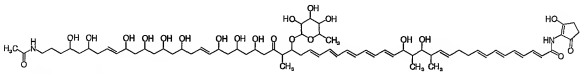
Compound 2(p)



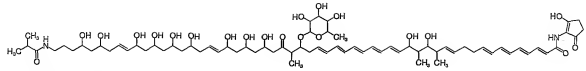
Compound 2(q)



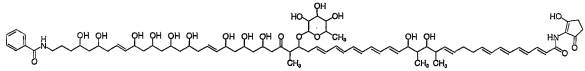
Compound 2(r)



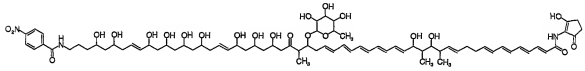
Compound 2(s)



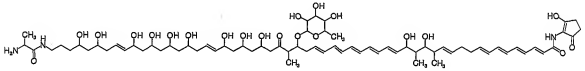
Compound 2(t)



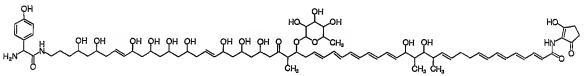
Compound 2(u)



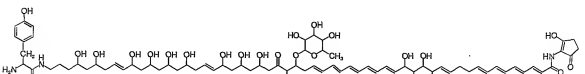
Compound 2(v)



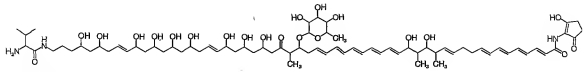
Compound 2(w)



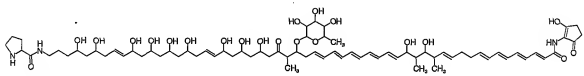
Compound 2(x)



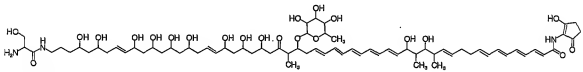
Compound 2(y)



Compound 2(z)



Compound 2(aa)



Compound 2(ab)

11. A method for producing the compound of claim 8, comprising the steps of cultivating cells derived from a *Streptomyces aizunensis* strain, incubating said cultured cells aerobically in a growth medium for such time as is required for production of said compound of claim 8, extracting said medium with a solvent and purifying the compound of claim 8 from the crude extract.

12. The method of claim 11 wherein said *Streptomyces aizunensis* strain is NRRL B-11277 or a mutant thereof.

13. The method of claim 12 wherein said mutant is strain [C03]023 (deposit accession number IDAC 070803-1) or [C03U03]023 (deposit accession number IDAC 231203-02).

14. The strain of *Streptomyces aizunensis* identified by deposit accession number IDAC 070803-1.

15. The strain of *Streptomyces aizunensis* identified by deposit accession number number IDAC 231203-02.

16. A pharmaceutical composition comprising a therapeutically effective amount of a compound of any one of claims 1 to 10, and a pharmaceutically acceptable carrier.

17. A pharmaceutical composition comprising a therapeutically effective amount of the compound of claim 8, and a pharmaceutically acceptable carrier.

18. A pharmaceutical composition comprising a therapeutically effective amount of a compound of claim 9, and a pharmaceutically acceptable carrier.

19. A method of treating a fungal infection in a mammal, comprising administering to said mammal suffering from said infection, a therapeutically effective amount of a compound of any one of claims 1 to 10.

20. The method of claim 19 wherein said fungal infection is caused by *Candida albicans*.

21. The method of claim 19 wherein said fungal infection is caused by a *Candida sp.*, wherein said *Candida sp.* is selected from the group consisting of *C. glabrata, C. lusitaniae C. parapsilosis, C. krusei* and *C. tropicalis*.

22. The method of claim 19 wherein said fungal infection is caused by an *Aspergillus sp.*, wherein said *Aspergillus sp.* is selected from the group consisting of *A. fumigatus, A. niger, A. terreus* and *A. flavus*.

23. The method of claim 19 wherein said fungal infection is caused by *Fusarium spp.; Scedosporium spp.; Cryptococcus spp.; Mucor ssp.; Histoplasma spp.; Trichosporon spp.; Blaspomyces spp.;* or *S. cerevisiae*.

24. A method of treating a fungal infection in a subject, comprising administering to said subject suffering from said infection, a therapeutically effective amount of a compound of any one of claims 1 to 10.

25. The method of claim 24 wherein said fungal infection is caused by a fungus selected from the group consisiting of *Candida albicans, Candida sp., Aspergillus sp., Fusarium spp.; Scedosporium spp.; Cryptococcus spp.; Mucor ssp.; Histoplasma spp.; Trichosporon spp.; Blaspomyces spp.;* and *S. cerevisiae.*

26. The method of claim 24 wherein said *Candida sp.* is selected from the group consisting of *C. glabrata, C. lusitaniae, C. parapsilosis, C. krusei* and *C. tropicalis.*

27. The method of claim 24 wherein said *Aspergillus sp.* is selected from the group consisting of *A. fumigatus, A. niger, A. terreus* and *A. flavus.*

28. A method of treating cancer in a subject, comprising administering to said subject suffering from said cancer, a therapeutically effective amount of a compound of any one of claims 1 to 10.

29. The method of claim 28, wherein said cancer is selected from the group consisting of leukemia, non-small cell lung cancer, colon cancer, CNS cancer, melanoma, ovarian cancer, renal cancer, prostate cancer and breast cancer.

Figure 1

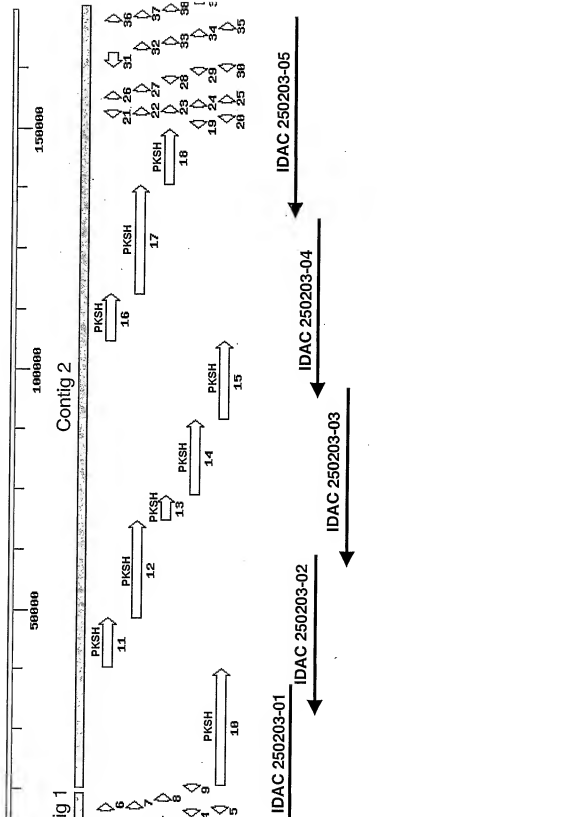

Gene Cluster for Production
of Compound 2(a)
Contig 1
Contig 2
0
50000
100000
150000
PKSH
IDAC 250203-01
IDAC 250203-02
IDAC 250203-03
IDAC 250203-04
IDAC 250203-05

Figure 2a

```
ORF10_pKS01          PIAIVGIGCHFPGGVQSPEALWNLVETGTDAISAFPTGRGWDLDALYDPDPDRAGTSYAR
ORF10_pKS02          PIVIVSMSCRFPGGVRTPEDLWQLLADGTDTVAAFPADRGWDLDGLYSADPERSGTSYTR
ORF10_pKS03          PIAIVSMSCRFPGGVRTPEDLWRLLVDGTDAVGAFPADRGWDLDRLYSPDPDQPGTSYTR
ORF10_pKS04          PIVIVGMGCRFPGGVRSPEDLWQLVATGGDGITGFPSDRGWNVEALYHPDPDHAGTSYTR
ORF10_pKS05          PIAIVAMSCRFPGGVRTPEDLWRLLSTGGDAIGEFPADRGWDLSRLYSPDPDKQGTFVAR
ORF11_pKS01          PIAIVAMSCRFPNGVGSPEDLWRLVDEGGDAIGPFPADRGWDIESLADPDPDRGTFFWT
ORF11_pKS02          PIAIVAMSCRYPGGVRTPEELWRLVETGGDAIAGLPGNRGWDTDALH---ADEDGRTFA-
ORF12_pKS01          PIAIVGMSCRFPGGVSSPEDLWRLVESGGDAISGFPVNRGWDIESLYDPDPDHEGTTYAR
ORF12_pKS02          PIAIVAMSCRFPGGIASPEDLWQLLVTGRDGITGFPADRGWDLDSLYSDDPDREGTSYAR
ORF12_pKS03          PIAIVSMSCRFPGGVRTPEDLWELLSTGGDAISDLPLDRGWDIDALYDADPSTQGTSYAR
ORF12_pKS04          PIAIVAMSCRFPGGVRTPEDLWQLLATGRDAIGEFPEDRGWDAEALFGP-QFEQDAFYAR
ORF13_pKS01          PIAITAMSCRFPGGVRSPEELWELLRTGGDAITAFPADRGWDLDNLFSDDPDDHNTSVTR
ORF14_pKS01          PIAIVGMGCRYPGGVTSPEELWQLVVDGGDAISGFPADRGWDMETVYHPDPEHPGTSYAN
ORF14_pKS02          PIAIVAMSCRFPGGVSEPELWQLLSTGRDAISGFPGDRGWDLDGLYDPEAGENTSYVR
ORF14_pKS03          PIAIVAMSCRYPGDVRTPEELWQLLTAGADGITRLPENRGWDMETGLYDPDPESQGTSYAR
ORF15_pKS01          PIAVVAMSCRYPGGIDTPEKLWDLVAHGRDAVSAYPTDRGWDAEVLFDPDPETGIEAYEQ
ORF15_pKS02          PIAIVAMSCRYPGGVTTPEELWQLLAGGGDAISGFPADRGWDVESLYDPDPDHPGTSYTR
ORF15_pKS03          PIAIVGLGCRYPGGVESPDDLWRLVLEGRDAITEFPEDRGWDVDALFDADPDQQGTSYAR
ORF16_pKS01          PIAIVAMSCRYPGGVRSPEDLWRLVENGDDAVSGFPVDRGWDVEALYDADPDSSGSSYVS
ORF16_pKS02          PIAIVAMSCRFPGGVRNPEELWQLLTSEGDGLSQFPLDRGWDVDALYDPNPDAQGTSYTR
ORF17_pKS01          PIAIVGMSCRFPGGIESPEGLWDLVAGGRDAITDFPTDRGWDIESLYDADPDQQGTSYTR
ORF17_pKS02          PIAIVGMSCRYPGGVTTPEELWQLVAGSVDAISPFPTDRGWNLDALYDADPDRAGTSYTR
ORF17_pKS03          PIAIVAMSCRFPGDVRTPEDLWELLAEGRDGISDLPDDRGWDTEALYDPDPSPGTSYAR
ORF17_pKS04          PIAIVGMSCRYPGGVETPEDLWRLVVGGGDAISEFPQGRGWDLESLYDPDPDKGGTSYTR
ORF18_pKS01          PIAIVGMSCRYPGDVESPEDLWRLVVSEETDAISPFPTDRGWDMGRLFDADPDGRGTSYVQ
ORF18_pKS02          PIAIVAMSCRFPGGVRSPEDLWGLVLDGRDAISDMPDDRGWDVEGLFDFDPDPRPGTSYSR
                     **.:.:.*::*..: .*: ** *:    * :   * .***:   : :
```

```
ORF10_pKS01          EGGFLHDADAFDAAFFGISPREALAMDPQQRLLLEASWEAFDRAGVDPAALRGGQVGVFV
ORF10_pKS02          EGGFLYDAADFDADFFGISPREALAMDPQQRLLLETAWETFERAGIDPASLRGSQAGVFV
ORF10_pKS03          EGGFFDGAADFDPGFFGISPREALAMDPQQRLLLETSWEAIERAGIDPSSLRGSQAGVFV
ORF10_pKS04          EGGFLHDAADFDPGFFGISPREALAMDPQQRLLLETSWEAFERAGIDPATLRGSRTGVFA
ORF10_pKS05          AGGFLYDAADFDADFFGISPREALAMDPQQRLLLETSWEAFERAGIDPSSLRGSQAGVFV
ORF11_pKS01          GGGFLDGATAFDPGFFGISPREALAMDPQQRQLLETSWEVFERAGIDPAAVRGSRTGVFV
ORF11_pKS02          -GGFLYDADSFDADFFGISPREALAMDPQQRLLLETSWEAIERAGIDPSSLRGSRAGVFV
ORF12_pKS01          DGGFLHEAADFDPAFFGISPREALAMDPQQRLLLETTWEVFERAGIDPASLRGSRAGVFV
ORF12_pKS02          EGGFLHEAAEFDASFFGISPREALAMDPQQRLLLETTWETFERAGIDPTSLRGSRTGVFV
ORF12_pKS03          AGGFLYDAADFDADFFGISPREALAMDPQQRLLLETSWEAFERAGIDPETLRGSQAGVFV
ORF12_pKS04          EGGFLYDVADFDPAFFGISPREALAMDPQQRLLLETSWEAFERAGIDPLSVRGSQAGVFV
ORF13_pKS01          EGGFLGEASSFDAAFFGISPREAMAMDPQQRLLLETSWEAFERAGIDPQALRGSQSGVFV
ORF14_pKS01          QGGFVRDFASFDPSLFGISPREALAMDPQQRLLLETSWEAFERAGIDPTSMRGKQVGVFV
ORF14_pKS02          EGGFLAGATEFDPAFFGISPREALAMDPQQRLLLETSWEAFERAGIDPATVRGEQIGVFT
ORF14_pKS03          DGGFLHDAAEFDASFFGISPREALAMDPQQRLLLETTWEVFERAGIAPSAVRGSRTGVFA
ORF15_pKS01          VGGFLHDAADFDPAFFGISPREALAMDPQQRLLLETSWEAFERAGIDPATLRGSRTGVFA
ORF15_pKS02          HGGFLRDAAAFDPTFFGISPREAVGTDPQQRLLLETTWEAFERAGIDPATVRGSRTGVFA
ORF15_pKS03          EGGFVRDAGHFDPAFFGISPREAVAMDPQQRLLLETSWEAFERAGIDPAALRGSRTGVFA
ORF16_pKS01          EGGFLYDAASFDPAPFGISPREALAMDPQQRLLLEASWEAFERAGIDPSSVRGSRTAVFA
ORF16_pKS02          EGGFLSDAAAFDSSFFGISPREALAMDPQQRLLLETSWEAFERAGIDPQTLRGSQSGVFV
ORF17_pKS01          EGGFLDGVGKFDASFFGISPRETLGMDPQQRLLLETSWEAFERAGIDAATLRGSKAGVFI
ORF17_pKS02          EGGFLHDAADFDPDVFGINPREALAMDPHQRLLLETSWEAFEQAGIAPSSMRGSRTGVFA
ORF17_pKS03          EGGFFYDAHHFDPAFFGINPREALAMDPQQRLLLETSWEAFERAGIDPTGLRGKQVGVFV
ORF17_pKS04          SGGFLHDAGRFDPAFFGISPREAVAMDPQQRLLLETSWEAFERAGIDPASMRGSRTGVFA
ORF18_pKS01          EGGFLHSANRFDPAFFGISPREAVAMDPQQRLLLETSWEAFERAGIDPTSLRGSRTGVFA
ORF18_pKS02          AGGFLHDAHHFDPTFFGISPREALATDPQQRLLLETSWEAFERAGIDPATVRGSRTGVFA
                      ***.     **.  ***.***:::  **:** ***::**.:::**: .  :** : .*:
```

Figure 2b

```
ORF10_pKS01          GAETQEYGPRLQ------DATDGFEGYLVTGNAASVASGRIAYTFGFEGPTVTVDTACSS
ORF10_pKS02          GTNGQDYLSLVTREGDG---LDGLEGHVGTGNAASVVSGRLSYVPGLEGPAITVDTACSS
ORF10_pKS03          GTNGQDYLSLITRESE------GLEGHLGTGNAGSVMSGRVSYVLGLEGPAVTVDTACSS
ORF10_pKS04          GVMYHDVTVTGIGDGGSAVELPEGVEGYLGTGNAGSIASGRIAYTFGLEGPAVTVDTACSS
ORF10_pKS05          GTNGQDYGAMLQTIPD------GIEGFLGTGNAASVVSGRLSYAPGLEGPAVTVDTACSA
ORF11_pKS01          GAGAMGYGADLKEA------PEGLEGLLLTGGATSVLSGRVSYVPGLEGPAATVDTACSS
ORF11_pKS02          GAAVSGYDAQLEQSG------VDGVLGHVMTGNAGSVMSGRVSYALGLEGPAVTVDTACSS
ORF12_pKS01          GASANNAYGAGSHDL------PDGVEGHLLTGTASSVLSGRLAVVPGLEGPAATIDTACSS
ORF12_pKS02          GSNAQDYLQLWLNDAD------GLEGHLGTGNAASVVSGRLSYTPGLEGPAVTVDTACSS
ORF12_pKS03          GTNGQDYLSVLLEEPE------GLEGHLGTGNAASVVSGRLSYVPGLEGPAVTVDTACSS
ORF12_pKS04          GTNGQDYLSLVLNSAD------GGDGFMSTGNSASVVSGRLSYVPGLEGPAVTVDTACSA
ORF13_pKS01          GINGSDYLTPLLEAAE------DYAGHLGTGNASSVMSGRLSYTPGLEGPAVTVDTACSA
ORF14_pKS01          GTSNHDYLSALLSS-------SENVEGYLGTGNAASVASGRLSYTPGLEGPAVTVDTACSS
ORF14_pKS02          GTNGQDYLNVILAAPD------GVEGFLGTGNAASVVSGRVSYVLGLEGPAVTVDTACSS
ORF14_pKS03          GVMYHDYGARLH------AVPEGVEGYLGTGSSSSIVSGRVAYTPGLEGPAVTVDTACSS
ORF15_pKS01          GLMYHDYAARLF------SVPEEIEGFLGNGSSGSIASGRIAYTLGLEGPAVTVDTACSS
ORF15_pKS02          GVMYHDYAALLE------RSKDGADGSLGSGSTGSIASGRVSYTFGLEGPAVTIDTACSS
ORF15_pKS03          GVMYHDYASRLT------ALPEGVEGFLGTGNAASVISGRLSYAPGLEGPAITVDTACSS
ORF16_pKS01          GVMYHDYTARLD------SVPEGVEGFLGTGSSGSIASGRVAYTPGLEGPAVTVDTACSS
ORF16_pKS02          GTNGSDYSNLVRAGAD------GLEGHLATGNAGSVVSGRLSYTPGLEGPAVTVDTACSA
ORF17_pKS01          GTNGQDYPELLREVPK------GVEGYLLTGNAASVVSGRISYTFGLEGPAVTVDTACSA
ORF17_pKS02          GVMYHDYLTRLP------AVPEGLEGYLGTGNAGSVASGRISYTFGLEGPAVTVDTACSS
ORF17_pKS03          GQMNNDYVSRLN------TVPEGVEGYLGTGGSSSIASGRVSYTFDFEGPAVTVDTACSS
ORF17_pKS04          GIMYHDYATRIT------SVPDGVEGYLGTGNSGSIASGRVSYAPGLEGPAVTVDTACSS
ORF18_pKS01          GVMYHDYASRLR------AVPEEVEGYLGTGGSSSIASGRVSYTPGLEGPALTVDTACSS
ORF18_pKS02          GVMYNDYGTLLH------RAPEGLEGYMGTSSGSVASGRVSYTFGLEGPAVTVDTACSS
                     *    *                    * : .. : *: ***::*...***: *:***M*:

ORF10_pKS01          SLAALHLAVQALRTGECSLALAGGVAVMASPGSFVSFSRQRGLAPDGRCKPFAAAADGTA
ORF10_pKS02          SLVALHLAVQALRQGECTLALAGGVTVMSTPDAFVDFSRQRGLAEDGRIKAFASAADGTG
ORF10_pKS03          SLVALHWAIQALRQGECSMALAGGVTVMSTPENFVDFSRQRGLAEDGRIKAFASAADGTG
ORF10_pKS04          SLVALHWAIQALRSGECTMALAGGVAVMATPETFVDFSRQRGLSADGRCKSFAAAADGTG
ORF10_pKS05          SLVALHWAVQALRSGECSLALAGGVTVMSSPGAYIDFSRQRGLAEDGRIKAFAAAADGTG
ORF11_pKS01          SLVALHLATQALRQGECSLALVGGVCVMPSPDVVFEFSRQRGLSPDGRCKSFAAGADGTG
ORF11_pKS02          SLVALHWAIQALRNGECSLALAGGVTVMSTPGTFSEFSQQGGLSPDGRCKAFASAADGTG
ORF12_pKS01          SSVALHMAVQALRQGECSLALAAGVTVLAGPDVFVEFSRQRGLSPDGRCRSFAESADGTG
ORF12_pKS02          SLVTLHLAAQALRRGECSMALAGAVTIMSTPGAFTEFSRQRGLAADGRIKAFAAAADGTS
ORF12_pKS03          SLVALHWAIQALRNGECSLALAGGVTVMSTPGTFIEFSRQRGLAEDGRIKAFAAAADGTG
ORF12_pKS04          SLVALHLAVQALRNGECSLALAGGVTVMSTPGAFAEFSRQRGLAEDGRIKAFAAAADGTG
ORF13_pKS01          SLVALHLAVQALRAGECSLAVAGGVHVMSTPGLFVEFSKQRGLSTDGRCKAFAAGADGFG
ORF14_pKS01          SSVALHLAVQALRNGECSLALAGGATLMSAPGTFIDYSQRGLATDGRCKAFSPDADGFS
ORF14_pKS02          SLVALHWAIQALRNGECTMALAGGVTVMSTPASFIDFSRQRGLAEDGRIKAFAAAADGTG
ORF14_pKS03          SLVALHLAAQALRNGECSLALAGGVTVMFTPGTFIEFSRQRGLAADGRCKSFAAAADGTG
ORF15_pKS01          SLVAVHLAAQALRNGECTLALAGGVTVMSTPGTFTEFSRQRGLSADGRCKSFAAAADGTG
ORF15_pKS02          SLVALHMAIQALRTGECDMALAGGVTVMATPGTFIGFSRQRGLSADGRCRAFSADADGTG
ORF15_pKS03          SLVALHLAVQALRNGECSLALAGGVTVMATPAAFVEFSRQRGLAADGRCKAFSAGADGTG
ORF16_pKS01          SLVTLHLAVQALRAGECSMALAGGVTVMATPATFTEFSRQRGLAPDGRCKFFAAAADGTG
ORF16_pKS02          SLVALHLAVQALRSGECSLALAGGVTVMSTPGTFIEFSRQRGLSTDGRCKAFSSDADGFS
ORF17_pKS01          SLVALHWAVQALRNDECSLALAGGVTVMSSPRAFVQFSRQRGLAPDGRCKPFADGADGTG
ORF17_pKS02          SLVALHLAAQALRNGECDMALAGGVTVMSTPDTFIDFSRQRGLSGNGRCKSFSADADGTG
ORF17_pKS03          SLVALHLAAQALRNGECTLALAGGVTIITPDVFTEFSRQRGLASDGRCKPFAEAADGTA
ORF17_pKS04          SLVALHWAIQALRNGECTMALAGGVTVMSTPGTFTEFSRQRGLAADGRIKSFAAAADGTS
ORF18_pKS01          SLVTLHLAMQALRKGECSLALAGGVTVMATPGTFTEFSRQRGLSPDGRCKSFADSADGTG
ORF18_pKS02          SLVTLHLAVQALRNGECDLALAGGVTVMATPGTFVAFSRQRGLASDGRCKPFAAAADGTA
                     * .::* * ****  ** :*:.... :: *  :  :*:* **: :** :.*:   ***
```

Figure 2c

```
ORF10_pKS01   WGEGVGMLLVERLSDARAKGHRILAVVRGSAINQDGASNGLTAPSGPSQQRVIRQALANA
ORF10_pKS02   WGEGVGMLLVERLSDARRNGHPVLAVVRGSAINQDGASNGLTAPNGPSQQRVIRQALAGA
ORF10_pKS03   WGEGVGMLLVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALASA
ORF10_pKS04   WAEGAGMLLVERLSDAERNGHPVLAVVRGSAINQDGASNGLTAPNGPSQQRVIREALASA
ORF10_pKS05   WGEGVGMLLVERLSDARRNGHPVLALVRGSAINQDGASNGLTAPNGPSQQRVIRQALANA
ORF11_pKS01   WSEGVGVLLVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPNGPAQQRVIRQALENA
ORF11_pKS02   WGEGVGMLLVERLSDARRNGHPVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRAALASA
ORF12_pKS01   WSEGAGVLLVERLSDARRNGHHILAVVRGSAVNQDGASNGLTAPNGPAQQKVIRQALESA
ORF12_pKS02   WSEGVGLLLVERLSDARRNGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIREALADA
ORF12_pKS03   WGEGVGMLLVERLSDAERNGHPVLAIVRGSAINQDGASNGLTAPNGPSQQRVIRAALASA
ORF12_pKS04   WGEGVGMLLVERLSDARRNGHPVLALVRGSAVNQDGASNGLTAPNGPSQQRVIRAALASA
ORF13_pKS01   PAEGVGVLLLERLSDARKNGRPVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANA
ORF14_pKS01   LAEGVGILLVERLSDARRKGHPVLAVVRGTAVNQDGASNGLTAPNGPSQQRVILQALSNA
ORF14_pKS02   WGEGVGILLVERLSDAQRNGHPVLAIVRGSAINQDGASNGLTAPNGPSQQRVIRQALASG
ORF14_pKS03   WGEGAGMLLLERLSDARRNGHQVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALANA
ORF15_pKS01   WGEGAGMLVLERLSEARRNGHPVLALVRGSAVNQDGASSGLTAPNGPSQQRVIRQALAGA
ORF15_pKS02   WGEGVGMLLVERLSDARRNGHPVLAVVRGSAINQDGASNGLTAPNGPSQQRVIRAALASA
ORF15_pKS03   WSEGAGVLLVERLSDARRNGHPVLAVVRGSAINQDGASNGLTAPNGPSQQRVIRQALASA
ORF16_pKS01   WGEGVGMLLVERLSDAQRNGHPILAVVRGSAINQDGASNGLTAPNGPSQQRVIHQALTNA
ORF16_pKS02   PAEGVGVLLVERLSDARRNGHPILAVVRGSAINQDGASNGLTAPNGPSQQRVIRQALANA
ORF17_pKS01   WGEGVGMLLVERLSDARRNGHPVLALVRGSAINQDGASNGLTAPNGPSQQRVIRQALTNA
ORF17_pKS02   WAEGAGMILVERLSDARRNGHQVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQALANA
ORF17_pKS03   WGEGVGMLLVERLSDARRNGHQVLAVVRGTAVNQDGASNGLTAPNGPSQQRVIRQALANA
ORF17_pKS04   WAEGAGMLLVERLSEARAKGHPVLAIVRGSAINQDGASNGLTAPNGPSQQRVIRQALAGA
ORF18_pKS01   WAEGAGMLLVERLSDARKNGHTVLAVVRGSAVNQDGASNGLTAPNGPSQQRVIRQALADA
ORF18_pKS02   WGEGVGMLLVERLSDARAKGHPVLAVVRGSAINQDGASNGLTAPNGPSQQRVIRQALASA
               .**.*::::****:*.  :*:  :**:***:*:******.*****.**:**:**   **  ..

ORF10_pKS01   GLSAAEVDVVEAHGTGTRLGDPIEAQALLATYGQEHTDDRPLWLGSLKSNIGHTQAAAGV
ORF10_pKS02   GLSAADVDAVEAHGTGTRLGDPIEAQALLATYGQGRPADRPLWLGSVKSNIGHTQAAAGV
ORF10_pKS03   GLSAADVDVVEAHGTGTKLGDPIEAQALLATYGQDRPAGRPLWLGSIKSNIGHTQAAAGV
ORF10_pKS04   DLSAADIDAVEAHGTGTRLGDPIEAQALLATYGREREAGRPLWLGSIKSNIGHTQAAAGV
ORF10_pKS05   GLSAAEVDAVEAHGTGTRLGDPIEVQALLATYGREREADQPLWLGSIKSNIGHTQAAAGV
ORF11_pKS01   RLSAAEVDVVEAHGTGTTLGDPIEAQALLATYGQDRPEGRPLRLGSLKSNIGHTQAAAGV
ORF11_pKS02   GLSAADVDVVEAHGTGTKLGDPIEAQALLATYGQDRPDGRPLWLGSIKSNIGHTQAAAGV
ORF12_pKS01   RLTPADIDAVEAHGTGTTLGDPIEAQALLATYGQGRTDGRPLWLGSLKSNLGHTQNAAGV
ORF12_pKS02   GLSAAEVDAVEAHGTGTTLGDPIEAQALLATYGQGRPDDQPLWLGSVKSNIGHTQAVAGA
ORF12_pKS03   GLSAADVDAVEAHGTGTTLGDPIEAQALLATYGQDRPADRPLQLGSIKSNIGHTQAAAGV
ORF12_pKS04   GLAPGDIDAVEAHGTGTKLGDPIEAQALLATYGQDRPADRPLQLGSIKSNIGHTQAAAGV
ORF13_pKS01   RLSTDQVDVVEAHGTGTSLGDPIEAQALIATYGQDRPADQPLLLGSVKSNIGHTQAAAGV
ORF14_pKS01   RLTPDQVDAVEAHGTGTGLGDPIEAQALIATYGQDRPDGRPLWLGSLKTNIGHAQAAAGV
ORF14_pKS02   GLTTMDVDAVEAHGTGTKLGDPIEAQALLATYGQDRPEGRPLLLGSIKSNLGHTQAAAGV
ORF14_pKS03   GVAAGHVDAVEAHGTGTTLGDPIEAQALLATYGQEHTDDRPLLLGSVKSNLGHTQAASGV
ORF15_pKS01   RLSATQVDAVEAHGTGTTLGDPIEAQALLATYGQDRPDGRPLWLGSIKSNMGHTQAAAGI
ORF15_pKS02   GLSAAEVDAVEAHGTGTTLGDPIEAQALLATYGREHTEDSPLWLGSIKSNMGHTQAAAGV
ORF15_pKS03   GLSAADVDVVEAHGTGTTLGDPIEAQALLATYGQEHTDEQPLLLGSIKSNFGHTQAAAGV
ORF16_pKS01   RLSAADVDVVEAHGTGTTLGDPIEAQALLATYGQDRPAGRPLLLGSIKSNIGHTQAAAGV
ORF16_pKS02   RLSAADVDVVEAHGTGTTLGDPIEAQALLATYGQDRPAGRPLLLGSIKSNIGHAQAAAGV
ORF17_pKS01   GLTPAQVDVVEAHGTGTTLGDPIEAQALLATYGQNRPEGRPLWLGSVKSNIGHTQAAAGV
ORF17_pKS02   GLTTAEVDVVEAHGTGTTLGDPIEAQALLATYGQDRPAGQPLRLGSIKSNIGHTQAAAGA
ORF17_pKS03   GLTAAEVDAVEAHGTGTRLGDPIEAQALLATYGQDRPEGSPLWLGSIKSNFGHTQAAAGV
ORF17_pKS04   RLTSDQIDVVEAHGTGTTLGDPIEAQALLATYGREREADQPLWLGSIKSNMGHTQAAAGV
ORF18_pKS01   RLTAADVDVVEAHGTGTTLGDPIEAQALLATYGREHTEDSPLWLGSVKSNLGHTQAAAGV
ORF18_pKS02   GLSAADVDVVEAHGTGTTLGDPIEAQALLATYGQEHTDDSPLWLGSIKSNFGHTQAAAGV
               ::.  .:*.*********  ******.***:****:  :     **  ***:*:*:**:*  .:*
```

Figure 2d

```
ORF10_pKS01          AGIIKMIMAMRHGVLPRTLHVDAPTPHVDWEAGAVTLLTEAVEWPESDRPRRAGVSSFGM
ORF10_pKS02          AGVMKMVMAMRHGVLPRTLHVDGPTPHVDWSAGDVALLTEQREWPATGHPRRAGVSSFGL
ORF10_pKS03          AGIIKMVLAMQHGVLPQTLHVDEPTPHVDWSAGEVTLLTEQTAWPTVDRPRRAGVSSFGI
ORF10_pKS04          AGIIKMVMAMRHGVLPQTLHVDEPSPQVDWEAGEVSLLTGAMPWPQTGRPRRAGVSSFGI
ORF10_pKS05          AGVIKMVLAMEHGVLPQTLHVDEPTPHVDWSAGDVALLTDAVEWPETGRPRRAGVSSFGF
ORF11_pKS01          AGIIKMVMAMRHGVLPQTLHVDEPTPNVDWTAGAVSLLTEPMPWPETGAPRRAAVSAFGV
ORF11_pKS02          AGIIKMVMAMRHGVLPRTLHVDEPTSHVDWSAGEVSLLSESAEWPLTERPRRAGVSSFGI
ORF12_pKS01          AGIIKMVMAMRHGVLPRTLHVDEPTSHVDWSTGAVALLTEPVEWPETGRPRRVGVSAFGV
ORF12_pKS02          AGIIKMVMAMRHGVLPQTLHIDEPTPYVDWSAGDIALLTEQRAWPETGRPRRAGVSSFGY
ORF12_pKS03          AGVIKMVLAMEHGVLPQSLHIDAPSPQVDWEAGDIALLTEQRQWPETGRPRRAGVSSFGF
ORF12_pKS04          AGLMKMVLAMQHGVLPQTLHVDEPTPHVDWSAGDIALLTERREWPETGRPRRAGISSFGV
ORF13_pKS01          AGVIKMVLAMQHGVLPQSLHIDEPSPHVDWESGAVSLLTEQTAWPETTHPRRAGVSSFGF
ORF14_pKS01          AGVIKSVMAMRHGVLPRTLHVDEPTPHVDWSAGDVSLLTEARPWPLGDQPRRIGVSSFGM
ORF14_pKS02          AGVMKMVLAMQHGVLPQTLHVDEPTPHVDWSAGDVALLADAVAWPETGRPRRAGVSSFGI
ORF14_pKS03          AGVIKMVMSMRHGVLPKTLHVDEPTPHVDWSAGAVSLLTEQTPWPETGRPRRAGVSSFGI
ORF15_pKS01          AGIIKMVMAMRHGILPKTLHVDEPTPNVDWSEGAVSLLTESVPWPETGAPRRAGVSSFGI
ORF15_pKS02          AGVIKMVLAIQHGVLPRTLHADRPSPHVDWSQGAVSLLTESVPWPETGRPRRAGVSSFGI
ORF15_pKS03          AGIIKIVQAMRHGVVPKTLHVDEPTPHVDWSAGAVSLLTEQVAWPETGRPRRAAISSFGF
ORF16_pKS01          ASIIKMVEAMRHGVVPKTLHLDEPTPHVDWEAGAVSLIGEKIAWPETGELRRAGVSSFGF
ORF16_pKS02          AGVMKMVLAMQHGVLPQSLHIAEPTPHVDWSAGEVALLTERRAWPETGRPRRAGVSSFGF
ORF17_pKS01          AGIIKMVLAMQHGVLPESLHIDQPSGNVDWAAGDVKLLTEAVPWPQTGQPRRAGVSSFGV
ORF17_pKS02          AGIIKMILAMRHGVMPPSLHIGEPSPHIDWTAGAVSLLTEAAEWPDAGRPRRAGISSFGV
ORF17_pKS03          AGIIKMVQAMHHGVLPKTLHVDAPSPHVDWSAGAVSLLTEQMAWPETGRPRRAGVSSFGM
ORF17_pKS04          AGIIKMIMAIRHGVLPKTLHVDEPTPHVDWEAGAVSLLTESVPWPETGRPRRAGVSSFGI
ORF18_pKS01          AGIIKMVMAIRHGIRPKTLHVDEPSTNVDWSAGAVSLLRESVEWPETGRPRRAAISSFGI
ORF18_pKS02          AGIIKMVQAMHHGVVPKTLHVDEPSPHVDWSAGAVSLLTEQMAWPETGRPRRAAISSFGI
                     *.::* : ::.** :* :**   *:  :**  * : *:      **       * .:*:**

ORF10_pKS01          SGTNAHVIVEEP
ORF10_pKS02          SGTNAHTIIEEA
ORF10_pKS03          SGTNAHTIIEQA
ORF10_pKS04          SGTNAHTIIEQP
ORF10_pKS05          SGTNAHTVLEQA
ORF11_pKS01          SGTNAHTIIEQA
ORF11_pKS02          SGTNAHTIIEQA
ORF12_pKS01          SGTNVHTIIEQA
ORF12_pKS02          SGTNAHAVIEQA
ORF12_pKS03          SGTNAHTIIEQA
ORF12_pKS04          SGTNAHTILEQA
ORF13_pKS01          SGTNAHVIVEQA
ORF14_pKS01          SGTNAHIILESA
ORF14_pKS02          SGTNAHTIIEQA
ORF14_pKS03          SGTNAHAIIEQA
ORF15_pKS01          SGTNAHTILEQA
ORF15_pKS02          SGTNAHTIIEQA
ORF15_pKS03          SGTNAHAIIEQA
ORF16_pKS01          SGTNAHVIVEQA
ORF16_pKS02          SGTNAHAIIEQA
ORF17_pKS01          SGTNAHTVIEQA
ORF17_pKS02          SGTNAHVIIEQP
ORF17_pKS03          SGTNAHAIIELA
ORF17_pKS04          SGTNAHTIIEQA
ORF18_pKS01          SGTNAHTIIEQA
ORF18_pKS02          SGTNAHTIIEQA
                     ****.* ::* .
```

Figure 3a

```
ORF10_pAT01      VLLPWALSAKTPEALRAQAR-------RLGTLIAAQP--HVTPL---DIGHSLATTRGRF
ORF10_pAT02      SVLPLLISAKSDAGLRAQSE-------QLATHLVGNP--DVPIG---DIAYSLTTGRSGL
ORF10_pAT03      LPLPYVLSAKSPEALRAQAS-------VLRTHLEATD--HNGPG-SDDLAFSLATARAHL
ORF10_pAT04      PVVPWVLSGKGEEALRAQAR-------QLQSYVLRAP--ELRPV---DIAGSLAVGRASF
ORF10_pAT05      GVMPWTLSAKSEAALRVQAE-------RLRTRIA-----SDPLLQPVDVAYSLATSRAAL
ORF11_pAT01      AVLPWTLSGRSTAALRAQAA-------RLLTTQGQDG--ATEPGRPLDIGYSLATTRAAL
ORF11_pAT02      GTVPYVLSAKSSDALRAQAR-------QLLAVVEAAE--SPRVA---DLAYSLATSRAGL
ORF12_pAT01      KALPWLLSAKGRDALRDRAA-------QLLAYAEEHP--DLRPV---DIAGSLAVGRPSF
ORF12_pAT02      PVLPLLVSGRTAPALRAQAERLRPAATALATGTVTNSG-ALEAL---DLGYSLATSRAAL
ORF12_pAT03      DFVPLMLSAKSDVALRAQAA-------SLRARLIAAP--DMRLS---DVGSTLTTGRSAF
ORF12_pAT04      AVSAWPLAGKTEAGLREQAE-------RLLAHIDAHS--ELR---PVDVGHSLATGRAAF
ORF13_pAT01      -VVPWVLSGKSAGALRAQAE-------RLSGFLAGASAVDVPSV---DVGWSLASSRAGL
ORF14_pAT01      GALPVVLSGRTEPALRAQAA-------ALHAHLAAHP--GLGIA---DLAFSQALTRAAL
ORF14_pAT02      GPQPWLLSAKTRDALHDQAR-------RLHAHAELNP--ELSPA---DLGLSLAAGRSAF
ORF14_pAT03      SSVPLIVSARGEDALRAQAR-------RLHAHVHADP--GLRAV---DLGLSLATTRSAL
ORF15_pAT01      VPPLWTLSAKSPAALRAQAG-------KLHAHLTAHP--GLRPG---DIAHSLAVGRTDF
ORF15_pAT02      SALPLQLAGRSAEALSAQAR-------ALSAHLT-----AHPDVPLADLAYSLATSRATF
ORF15_pAT03      GSLPWLLSAKGADALRDQAA-------RLRAHAIGHP--ELSLA---DIGYALATSRTAL
ORF16_pAT01      -VVPWVLSGKSAGALRAQAE-------RLSGWLAGASAAGVASV---DVGWSLASSRAGL
ORF16_pAT02      -VLPWTLTAKTEKALQGQAE-------RLLTQLT--TRSDLRLV---DVGHSLATTRTAL
ORF17_pAT01      GPVPVLVSGQSDAALRAQAE-------RLAAHLRAHPGLGADTGTLTDLGFSLATSRSSL
ORF17_pAT02      TPLPFALSGRTPAALRAQAA-------RLIGHLAPRP--EAAPA---DVALSLATTRTAL
ORF17_pAT03      AALPWNLSARTPDALRAQGE-------RLLSHLETHCE-THPETVLADIGHSLTTGRALF
ORF17_pAT04      RVLPFVLSAKSAGALRGQAV-------RLKAHVEASP--EVSGAGAVDVAYSLATRRAVF
ORF18_pAT01      PVLPWPVSARTEEALHAQAE-------RLLAHVR-----TNPDQAPVGVALSLATGRAAL
ORF18_pAT02      RVLPFVLSAKSAGALRGQAV-------RLKAHVEASP--EVSGAGAADVAYSLATRRAVF
                       ::.:   .*  :.         *                 .:.  : :  *  :

ORF10_pAT01      EQRAIVLGDD---REAFLDALHALAEGN--------DTPSVVQGAA-APGKLAFLFTGQG
ORF10_pAT02      ETRAILVGDADN-RTGLAAALRSLAAGE--------QAPGLVQGTV-TEGGLAFLFTGQG
ORF10_pAT03      EHRAVLT-ADD--PQEFREALARLADGD--------PSPRITTGAV-SDGRTAFLFTGQG
ORF10_pAT04      EDRAAVV--AAD-REGLLAALAALADGG--------SATGAVEGSA-VGGKLAFLFTGQG
ORF10_pAT05      ERRAVVVATE---RDEFLAGLKALASGQ--------PAPGLVQGR-VTEGGLAFLFTGQG
ORF11_pAT01      EHRAVLL-GRT--EDDFAAALSALAEGA--------ESAGLVQGRV-TEGGLAFLFTGQG
ORF11_pAT02      DHRAALVADD---RENLTRALAALAADE--------QVPGLVRGTA-TGGGLAFLFTGQG
ORF12_pAT01      EDRAAVV-AAD--REGLLAGLAALADGG--------SATGLVKGSSQLVGKLAFLFTGQG
ORF12_pAT02      EHRAVLIGTPSD-GQALASRLDALAAGE--------QVPGLVQGTA-SGGGLAFLFTGQG
ORF12_pAT03      ERRAALV--AGG-REGLLAGLEALADGG--------SAAGLVEGSP-VSGKLAFLFTGQG
ORF12_pAT04      DHRAVLVAGDD--RSEFRRALAALASGE--------SVAQVVQGIARPDQQVAFLFTGQG
ORF13_pAT01      EHRAVVLG-------DHAAGVAAVASGV--------MAAGVVTGSV-VGGKTAFVFPGQG
ORF14_pAT01      DRRAAVVADD---RDALLAGLAALAEGR--------PSADVVEGSA-TDGKLAFLFTGQG
ORF14_pAT02      ERRAAVI-AAD--RDGLLAGLAALADGG--------AAAGLVEGSP-VAGKLAFLFTGQG
ORF14_pAT03      EQRAALVAGD---RAELLRGLDALARGE--------DTAGLVRGTA-REGQVAFLFTGQG
ORF15_pAT01      EHRAVLT-SADG-PVGLVRALEALADSAPEDTAPADRAPGVTRGRP-VAGKLAFLFTGQG
ORF15_pAT02      DHRAVLVATEGTTAATAVTALDALADRR--------TAPGLVRGTASKGGRTAFLFTGQG
ORF15_pAT03      DRRAAVVAGD---REEFLAGLAALAEGA--------TAAGLTEGSP-AGGKLAFLFTGQG
ORF16_pAT01      EHRAVVLG-------DHAAGVGAVASGV--------MAAGVVTGSV-VGGKTAFVFPGQG
ORF16_pAT02      DQRAVLIGRDRP---DYLGALTALAAGD--------TSPLLVQGAV-VGGKTAFVFPGQG
ORF17_pAT01      DRRAVLFGDR----DSLLADLSALAEGE--------QPAGPVLGAV-GEGKTAFLFTGQG
ORF17_pAT02      DRRAAVIAHD---RTELLAGLTALAEGH--------DSARLVQHTA-ADGRTAILFTGQG
ORF17_pAT03      EHRATVVAGD---RDGFRAGLAALAEGR--------TAAGLIQGSSSTGGRTAFLFTGQG
ORF17_pAT04      DHRAVVVAGD---REELLRSLAAVESEG--------AAAGVTRGAVG-GGKLAFLFTGQG
ORF18_pAT01      EHRAVVVATD---RETALADLAALASGE--------TSARVVLGEPGARGKTAFLFTGQG
ORF18_pAT02      DHRAVVVAGD---REELLRALAAVESEG--------TAAGVTRGTAG-GGKLAFLFTGQG
                 : ** :              :  :              .             *::*.***
```

Figure 3b

```
                                                                                 M1
ORF10_pAT01    SQRLGMGRELYETHPVFADALDDACWYLDDQLELPL--LDVLFADEGSPEAALLHQTAYT
ORF10_pAT02    SQRLGMGRELYETYPVFADALDAVCARMD--LEVPL--RDVLFGAY----AGLLDETAYT
ORF10_pAT03    SQRLGMGRELYEAYPVFADALDAVCAHVDAHLEVPL--KDVLFGAD----AGLLDQTAYT
ORF10_pAT04    SQRLGMGRELYEAYPVFAEALDAVCAR----LELPL--KDVLFGAD----AGLLDETAYT
ORF10_pAT05    SQRLGMGRELYETYPVFADALDAVCVR----LELPL--MDVLFGTE----RDALDETGYT
ORF11_pAT01    SQRLGMGRELYEAYPVFADALDAVCAR----LELPL--KDVLFGAD----AGLLDETAYT
ORF11_pAT02    SQRLGMGRELYETYPVFARALDAV----DARLELPM--KEVLFGAD----ADLLNETAHT
ORF12_pAT01    SQRLGMGRELYETYPVFAQALDAVCER----LELPL--KNVLFGTD----SAALDETSYT
ORF12_pAT02    SQRLGMGRELYETYPVFAEALDAVCAR----LELPL--KEVLFGAD----GAALDQTAVT
ORF12_pAT03    SQRLGMGRELYEAYPVFADALDAVCVR----LELPL--MDVLFGAD----AGLLNETAYT
ORF12_pAT04    SQRLGMGRELYETYPVFADALDAVCAR----LELPL--KDVLFGGD----ADRLNETAYT
ORF13_pAT01    SQWVGMAVGLLDSSPVFAARVEECAKALEPFTDWSL--VDVLRGVEG---APSLERVDVV
ORF14_pAT01    SQRPGMGRELYATYPVFAQALDAVCER----LELPL--KDVLFGTDGAA-GAALDETAYT
ORF14_pAT02    SQRLGMGRELYDTYPVFADALDAVCAHVDAHLEVPL--KDVLFGAD----TGLLDQTAYT
ORF14_pAT03    SQRPGMGRELYDAHPVFADALDEICGELDRHLEVPL--KGVLFATE----GDLIHQTAYT
ORF15_pAT01    SQRLGMGRELYETYPVFAQALDAVCERLN--LEVPL--RDVLFGAD----AGLLDQTVYT
ORF15_pAT02    SQRLGMGRELYEAHPVFARALDAVCDR----LELPL--KDVLFGTD----AGLLNETVYT
ORF15_pAT03    SQRLAMGRELYSAHPVFARALDAVCDG--LALDVPL--KQVLFGSD----ADLLDRTAYT
ORF16_pAT01    SQWVGMAVGLLDSSPVFAARVDECAKALEPFTDWSL--VDVLRGVEG---APSLERVDVV
ORF16_pAT02    SQWVGMAVALLDASPVFAARVDECAKALEPFTDWSL--RDVLRGVTG---APSLDRVDVV
ORF17_pAT01    SQRLGMGRELYATHPGFARALDEVRAELDQHLERPL--FDVLFAAEGTPEADLLDETAYT
ORF17_pAT02    SQRPGMGRELYETYPAFAEALDAVCAELDPHLEQPL--KEVLFTAD----GDLLNRTGRT
ORF17_pAT03    SQRLGMGRELYEAYPVFARALDEVCAR----LELPLPLKDVLFGTD----TGLLNETAYT
ORF17_pAT04    SQRLGMGRELYETYPVFARALDAACAR----LELPL--KDALFGTD----AGLLGETAYT
ORF18_pAT01    SQRLGMGRELYEEYPVFADALDAVCAR----LELPL--KDVLFGAD----ARLLDETAYT
ORF18_pAT02    SQRLGMGRELYETYPVFARALDAACAG----LELPL--KDALFGAD----AGLLDETAYT
               **   .*.  *    * **  ::          : .:    .*             : ..  .

                       M2          M3
ORF10_pAT01    QPALFAVEVALFRLVDSWGLKPDFVAGHSIGEIAAAHVAGVFSLEDACMLVAARGRLMQA
ORF10_pAT02    QPALFAVEVALFRLVESWGLRPDFVAGHSIGEIAAAHVAGVLSLDDACALVEARGRLMGA
ORF10_pAT03    QPALFAVEVALFRLVESWGVKPDFVAGHSIGEIAAAHVAGVFSLQDASELVFARGRLMQA
ORF10_pAT04    QPALFAVEVALFRLVESWGLRPDFVAGHSIGEIAAAHVAGVFSLDDACALVEARGRLMGA
ORF10_pAT05    QPALFAVEVALFRLVESWGVRPDFLAGHSIGEIAAAHVAGVFSLDDACALVEARGRLMQA
ORF11_pAT01    QPALFAVEVALFRLVESWGVKPDFVAGHSIGEIAAAHVAGVFSLEDACALVSARGRLMGA
ORF11_pAT02    QPALFAVEVALFRLLESWGVRPDVLAGHSIGEIAAAHVAGVFSLDDACTLVEARGRLMQA
ORF12_pAT01    QPALFAVEVALFRLVESWGLKPDFLAGHSIGEIAAAHVAGVFSLDDACALVSARGRLMGA
ORF12_pAT02    QPALFAIEVALFRLVESWGLRPDFVAGHSIGEIAAAHVAGVFSLEDACRLVEARGRLMQA
ORF12_pAT03    QPALFAVEVALFRLVESWGLRPDFLAGHSIGEIAAAHVAGVLSLDDACALVEARGRLMGA
ORF12_pAT04    QPALFAVEVALFRLVESWGVRPDFLAGHSIGEIAAAHVAGVFSLDDACALVEARGRLMQA
ORF13_pAT01    QPALFAVMVSLAEVWRAAGVRPGAVIGHSQGEIAAACVAGILSLEDAARVVALRSQAIGR
ORF14_pAT01    QPALFAVEVALFRLVESWGLKPDYLAGHSIGEIAAAHVAGVFSLEDACTLVEARGRLMQA
ORF14_pAT02    QPALFAVEVALFRLVESWGLRPDFLAGHSIGEIAAAHVAGVFSLQDASELVVARGRLMQA
ORF14_pAT03    QPALFAVEVALFRLLESRGVQPDFLAGHSIGEIAAAHVAGVFSLQDASELVAARGRLMQA
ORF15_pAT01    QTALFAVEVALFRLVESWGLKPDFLAGHSIGEIAAAHVAGVFSLEDACALVSARGRLMGA
ORF15_pAT02    QPGLFAVEVALFRLLESWGVKPDFLAGHSIGEIAAAHVAGVLSLDDVCALVEARGRLMGA
ORF15_pAT03    QPALFAVEVALFRLVESWGLKPDFLAGHSIGEITAAHVAGVLSLDDACTLVAARGRLMQA
ORF16_pAT01    QPALFAVMVSLAEVWRAAGVRPGAVIGHSQGEIAAACVAGILSLEDAARVVALRSQAIGR
ORF16_pAT02    QPALFAVMVSLAEVWRAAGVRPDAVIGHSQGEIAAACVAGILSLEDAARVVALRSQAIGR
ORF17_pAT01    QSALFAVEVALFRQLEQWGVGADFLIGHSIGELAAAHVSGVFTLADAAKLVAARGRLMQA
ORF17_pAT02    QPALFALETALYRLVESWGVRPDFVAGHSIGEITAAHVAGVLSLPDAATLVAARGRLMQE
ORF17_pAT03    QPALFAVEVALFRLVESWGLKPDFLAGHSIGEIAAAHVAGVLSLEDACALVSARGRLMGA
ORF17_pAT04    QPALFAVEVALFRLLESWGVRPDFLAGHSIGEIAAAHVAGVLSLDDACALVEARGRLMQA
ORF18_pAT01    QPALFAVEVALFRLVESWGLKPDFLAGHSIGEIAAAHVAGVFSLEDACALVSARGRLMGA
ORF18_pAT02    QPALFAVEVALFRLLESWGVRPDFLAGHSIGEIAAAHVAGVLSLDDACALVAARGRLMQA
               *..***: .:* .      *: .. : *** **::** *:*:::* *.. :*  *.: :
```

Figure 3c

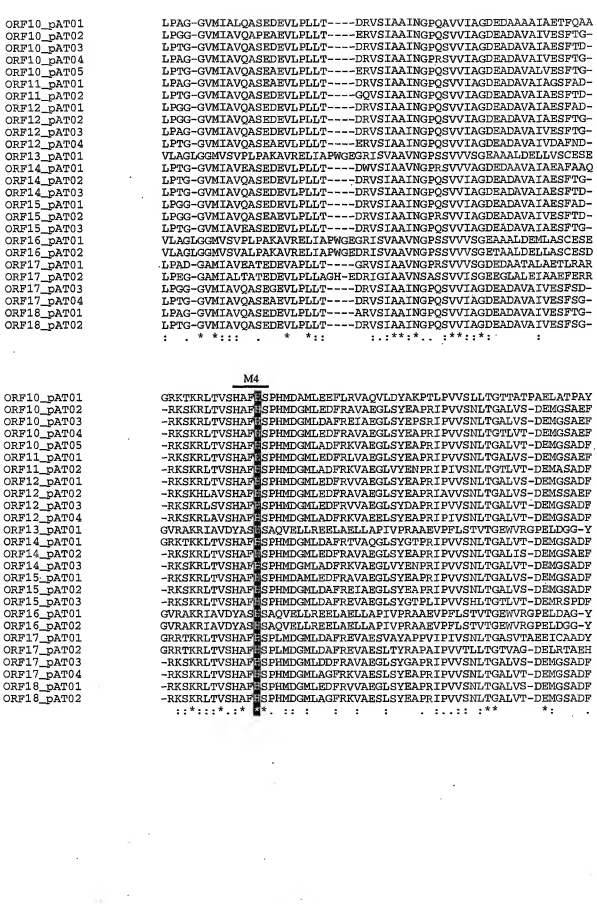

ORF10_pAT01 LPAG-GVMIALQASEDEVLPLLT----DRVSIAAINGPQAVVIAGDEDAAAAIAETFQAA
ORF10_pAT02 LPGG-GVMIAVQAPEAEVLPLLT----ERVSIAAINGPQSVVIAGDEADAVAIVESFTG-
ORF10_pAT03 LPTG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFTD-
ORF10_pAT04 LPAG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPRSVVIAGDEADAVAIVESFTG-
ORF10_pAT05 LPTG-GVMIAVQASEAEVLPLLT----ERVSIAAINGPQSVVIAGDEADAVAIVESFTG-
ORF11_pAT01 LPAG-GVMIAVQASEAEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAGSFAD-
ORF11_pAT02 LPTG-GVMIAVQASEDEVLPLLT----GQVSIAAINGPQSVVIAGDEADAVAIAESFTD-
ORF12_pAT01 LPGG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFAD-
ORF12_pAT02 LPGG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFTG-
ORF12_pAT03 LPAG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIVESFTG-
ORF12_pAT04 LPTG-GVMIAVQASEAEVLPLLT----ERVSIAAINGPQSVVIAGDEADAVAIVDAFND-
ORF13_pAT01 VLAGLGGMVSVPLPAKAVRELIAPWGEGRISVAAVNGPSSVVVSGEAAALDELLVSCESE
ORF14_pAT01 LPTG-GVMIAVQASEDEVLPLLT----DWVSIAAVNGPRSVVVAGDEDAAVAIAEAFAAQ
ORF14_pAT02 LPTG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFTG-
ORF14_pAT03 LPTG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFTD-
ORF15_pAT01 LPGG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFAD-
ORF15_pAT02 LPGG-GVMIAVQASEAEVLPLLT----DRVSIAAINGPRSVVIAGDEADAVAIVESFTD-
ORF15_pAT03 LPTG-GVMIAVEASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIAESFTG-
ORF16_pAT01 VLAGLGGMVSVPLPAKAVRELIAPWGEGRISVAAVNGPSSVVVSGEAAALDEMLASCESE
ORF16_pAT02 VLAGLGGMVSVALPAKAVRELIAPWGEDRISVAAVNGPSSVVVSGEPAALDELLASCESD
ORF17_pAT01 LPAD-GAMIAVEATSDEVAPLLT----GRVSIAAVNGPRSVVVSGDEDAATALAETLRAR
ORF17_pAT02 LPEG-GAMIALTATEDEVLPLLAGH-EDRIGIAAVNSASSVVISGEEGLALEIAAEFERR
ORF17_pAT03 LPGG-GVMIAVQASEGEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIVESFSD-
ORF17_pAT04 LPTG-GVMIAVQASEAEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIVESFSG-
ORF18_pAT01 LPAG-GVMIAVQASEDEVLPLLT----ARVSIAAINGPQSVVIAGDEADAVAIVESFTG-
ORF18_pAT02 LPTG-GVMIAVQASEDEVLPLLT----DRVSIAAINGPQSVVIAGDEADAVAIVESFSG-
M4
ORF10_pAT01 GRKTKRLTVSHAPFSPHMDGMLEEFLRVAQVLDYAKPTLPVVSLLTGTTATPAELATPAY
ORF10_pAT02 -RKSKRLTVSHAPFSPHMDGMLEDFRAVAEGLSYEAPRIPVVSNLTGALVS-DEMGSAEF
ORF10_pAT03 -RKSKRLTVSHAPFSPHMDGMLDAFREIAEGLSYEPSRIPVVSNLTGALVS-DEMGSAEF
ORF10_pAT04 -RKSKRLTVSHAPFSPHMDGMLEDFRAVAEGLSYEAPRIPVVSNLTGTLVT-DEMGSAEF
ORF10_pAT05 -RKSKRLTVSHAPFSPHMDGMLADFRKVAEGLSYEAPRIPVVSNLTGALVT-DEMGSADF
ORF11_pAT01 -RKSKRLTVSHAPFSPHMDGMLEDFRLVAEGLSYEAPRIPVVSNLTGALVS-DEMGSAEF
ORF11_pAT02 -RKSKRLTVSHAPFSPHMDGMLADFRKVAEGLVYENPRIPIVSNLTGTLVT-DEMASADF
ORF12_pAT01 -RKSKRLTVSHAPFSPHMDGMLEDFRVVAEGLSYEAPRIPVVSNLTGALVS-DEMGSADF
ORF12_pAT02 -RKSKHLAVSHAPFSPHMDGMLEDFRAVAEGLSYEAPRIAVVSNLTGALVS-DEMSSAEF
ORF12_pAT03 -RKSKRLSVSHAPFSPHMDGMLEDFRVVAEGLSYDAPRIPVVSNLTGALVT-DEMGSADF
ORF12_pAT04 -RKSKRLAVSHAPFSPHMDGMLADFRKVAEGLSYEAPRIPIVSNLTGALVT-DEMGSADF
ORF13_pAT01 GVRAKRIAVDYASFSAQVELLRREELAELLAPIVPRAAEVPFLSTVTGEWVRGPELDGG-Y
ORF14_pAT01 GRKTKKLTVSHAPFSPHMDGMLDAFRTVAQGLSYGTPRIPVVSNLTGALVT-DEMGSADF
ORF14_pAT02 -RKSKRLTVSHAPFSPHMDGMLEDFRAVAEGLSYEAPRIPVVSNLTGALIS-DEMGSAEF
ORF14_pAT03 -RKSKRLTVSHAPFSPHMDGMLADFRKVAEGLVYENPRIPVVSNLTGALVT-DEMGSADF
ORF15_pAT01 -RKSKRLTVSHAPFSPHMDAMLEDFRAVAEGLSYEAPRIPVVSNLTGALVS-DEMGSADF
ORF15_pAT02 -RKSKRLTVSHAPFSPHMDGMLDAFRIAEGLSYEAPRIPVVSNLTGALVS-DEMGSADF
ORF15_pAT03 -RKSKRLTVSHAPFSPHMDGMLDAFREVAEGLSYGTPLIPVVSNLTGTLVT-DEMGSPDF
ORF16_pAT01 GVRAKRIAVDYASFSAQVELLRREELAELLAPIVPRAAEVPFLSTVTGEWVRGPELDAG-Y
ORF16_pAT02 GVRAKRIAVDYASFSAQVELLRREELAELLAPIVPRAAEVPFLSTVTGEWVRGPELDGG-Y
ORF17_pAT01 GRRTKRLTVSHAPFSPLMDGMLDAFREVAESLVYAPAPIVPIVSNLTGASVTAEEICAADY
ORF17_pAT02 GRRTKRLTVSHAPFSPLMDGMLDAFREVAESLTYRAPAIPVVTLLTGTVAG-DELRTAEH
ORF17_pAT03 -RKSKRLTVSHAPFSPHMDGMLDDFRAVAEGLSYGAPRIPVVSNLTGALVS-DEMGSADF
ORF17_pAT04 -RKSKRLTVSHAPFSPHMDGMLAGFRKVAESLSYEAPRIPVVSNLTGALVT-DEMGSADF
ORF18_pAT01 -RKSKRLTVSHAPFSPHMDGMLEDFRVVAEGLSYEAPRIPVVSNLTGALVS-DEMGSADF
ORF18_pAT02 -RKSKRLTVSHAPFSPHMDGMLAGFRKVAESLSYEAPRIPVVSNLTGALVT-DEMGSADF

Figure 3d

```
ORF10_pAT01     WVRHVRDAVRYLDGVRTLHQRGVRTFLELGPDAVLTAMAQDCVDP-----QGAAFAPALR
ORF10_pAT02     WVRHVREAVRFLDGMRVLEAAGVTTYVELGPGGVLSALAQECVSG-----DGAAFVPVLR
ORF10_pAT03     WVRHVREAVRFLDGIRTLEAAGVTKYVELGPDGVLSAMAQDCVSG-----EGSVFIPVLR
ORF10_pAT04     WVRHVREAVRFLDGIRALEAAGVTTYVELGPGGVLSALAQECVSG-----DGAAFVPVLR
ORF10_pAT05     WVRHVREAVRFLDGTRTLEALGVTTYVELGPDGVLSAMAQECVTG-----EDSVFVPVLR
ORF11_pAT01     WVRHVREAVRFLDGIRTLEAAGVTKYVELGPDGVLSAMAQDCVSG-----EGSVFIPVLR
ORF11_pAT02     WVRHVREAVRFLDGIRALESRGVTTYIELGPDGVLSALAQDCLTAGTG-TGTAIFAPVLR
ORF12_pAT01     WVRHVRETVRFLDGIRTLEAAGVTKYVELGPDGVLSALAQDCVSG-----EDSVFIPVLR
ORF12_pAT02     WVRHVREAVRFLDGIRALEAAGVTTYVELGPGGVLSALAQECVSG-----DGAAFVPVLR
ORF12_pAT03     WVRHVREAVRFLDGIRALEAAGVTTYVELGPDGVLSAMAQECVTE-----GGAAFVPVLR
ORF12_pAT04     WVRHVREAVRFLDGIRALEAAGVTVYVELGPDGVLSAMAQECVTG-----EGAAFVPALR
ORF13_pAT01     WFQNLRRTVELEEATRTLLEQGFGVFVESSPHPVLSVGMQETVEDAG---REAAVLGSLR
ORF14_pAT01     WVRHVREAVRFLDGIRWLESRGVTTYIELGPGGVLSALGQDCQTATG--PRAAAFLPALR
ORF14_pAT02     WVRHVREAVRFLDGIRTLEAAGVTKYVELGPDGVLSAMAQDCVSG-----EGSVFIPVLR
ORF14_pAT03     WVRHVREAVRFLDGIRALEAAGVTTHIELGPDGVLCAMAQECVSG-----EDTVFVPVLR
ORF15_pAT01     WVRHVRETVRFLDGIRALTERNVVHFVELGPDAVLSAMAQDCPSA-----DTAAFVPVLR
ORF15_pAT02     WVRHVREAVRFLDGIHALEAAGVTTYVELGPDGVLSAMAQECVTG-----EDSVFVPVLR
ORF15_pAT03     WVRHVREAVRFLDGIRTLEDAGVTTYIELGPGGVLSAMGQSCVTR-----DDAAFLPALR
ORF16_pAT01     WFQNLRRTVELEEATRTLLEQGFGVFVESSPHPVLSVGMQETVEDAG---REAAVLGSLR
ORF16_pAT02     WFQNLRRTVELEEATRTLLEQGFGVFVESSPHPVLTMGVQETVEDAG---RDAAVLGSLR
ORF17_pAT01     WVRHVREAVRFLDGVRKLSAQGVTTFVEVGPGGVLTALAQECVTG-----QDAVFVPVLR
ORF17_pAT02     WVSHVREAVRFLDGIRTLDAEHVTTYLELGPQGVLSGLGRDCLTDPADPADTAVFVPALR
ORF17_pAT03     WVRHVREAVRFLDGIRALEAAGVTTYIELGPDGILSAMAQECITG-----EGAAFAPVLR
ORF17_pAT04     WVRHVREAVRFLDGIRTLEAAGVATYVELGPDGVLSAMAQDCVTG-----EGAAFAPALR
ORF18_pAT01     WVRHVREAVRFLDGIRALEAAGVTTYVELGPDGVLSAMAQACVTG-----ENSVFVPVLR
ORF18_pAT02     WVRHVREAVRFLDGIRALEAAGVTAYVELGPDGVLSALAQECVTG-----EGAAFAPALR
                *. ::* :*.  :. : *    .  .:* .*  :*    :            : .   **


ORF10_pAT01     SGRPEAATVLNAVAHAHVRGAETDWAAFFAGTGAQRVDLPTYAFQRQRYWM--
ORF10_pAT02     SGRPEAETAVTALAQAHVRGVDVDWAAFFSGTGVQRVDLPTYAFQRQRFWP--
ORF10_pAT03     KARPEAESVTTALASAHVHGIPVDWQAYFAGTGAQRVDLPTYAFQRQRYWP--
ORF10_pAT04     SGRSEAETAVTALAQAHVRGVNVDWAAFFAGTGAERVDLPTYAFQRQRYWL--
ORF10_pAT05     SGRPEAESVTTALAQVHVRGIAVDWQAYFAGTGAQRVDLPTYAFQRRRYWL--
ORF11_pAT01     KARPEAESVTTALATAHVHGIPVDWQAFYAGTGAQRVDLPTYAFQHERYWL--
ORF11_pAT02     AARPEAESVTTALATAHVHGTPVDWRAYFAGTGARRADLPTYPFQGRRYWP--
ORF12_pAT01     KARPEAETVATALASAHVHGIPVDWRAYFAGTGAQRVDLPTYPFQRQRYWI--
ORF12_pAT02     SGRSEAETVVTALAQAHVRGVEVDWAAFFAGTGAERIDLPTYAFQRQRYWP--
ORF12_pAT03     KGRPEAETVMATLGQAHVRGVAVDWHSVYG-TGAQRVDLPTYSFQRQRYWP--
ORF12_pAT04     KGRPEAETITAALAHAHTHGIAVDWQAYFAGTGAQRVDLPTYAFQRQRYWV--
ORF13_pAT01     RGEGGLERFWLSLGEAWVRGVGVDWHAVFAGTGAQRVDLPTYAFQSQRFWPEA
ORF14_pAT01     TGRPEASSLTAAVAGAHVRGLSPDWTVRFAGTGAQRVELPTYAFQRELYWP--
ORF14_pAT02     KARPEPESVTTALTTAHVHGIPVDWQAFFAGTGARRVDLPTYAFQRQRYWP--
ORF14_pAT03     PGRPEAETVTTALARVHVQGVPVDWQAYFSGTGAQRVDLPTYAFQRKRYWL--
ORF15_pAT01     KGRSETGSLTDALARLHVGGVAVDWDAYYSGTDVQRVDLPTYAFQRAHYWL--
ORF15_pAT02     SGRPEAESVTTALAQAHVRGIAVDWQAYFAGTSAQRVDLPTYRFQREHYWP--
ORF15_pAT03     ADRSEEETLTSAVARAHLRGITVDWDAYYSGTGARRVDLPTYAFQRQRYWL--
ORF16_pAT01     RGEGGLERFWLSLGEAWVRGVAVDWHAVFAGTGARRVDLPTYAFQQEHYWLES
ORF16_pAT02     RGEGGLERFWLSLGEAWVRGVGVDWSAVFAGTGARRVDLPTYAFQSQRFWPEA
ORF17_pAT01     GDRPEAAAFATAVAQAHVHGVAVDWSAVFAGRGATRIDLPTYAFQRELYWP--
ORF17_pAT02     RDRGEAEALTAAIAAAHTRGVPLDWSAYFAGTGARRVELPTYAFQRERFWL--
ORF17_pAT03     AGRDEAETVLSALAAAHVRGVPVDWQAFYAPAGAQRVPLPTYAFQRSVYWL--
ORF17_pAT04     KGRPETETITTALALAHAHGTSVDWETYFAGTGAQGVELPTYAFQRDWYWL--
ORF18_pAT01     SGRSEAESVTTALAQAHVRGIAVDWQAYFAGTGAERVDLPTYAFQRDHYWL--
ORF18_pAT02     KGRPEAETITTALALAHNHGTSVDWETYFSGTGAQRVDLPTYAFQRERYWI--
                  .        ::      *   **   :.  ..    **** **   :*
```

Figure 4

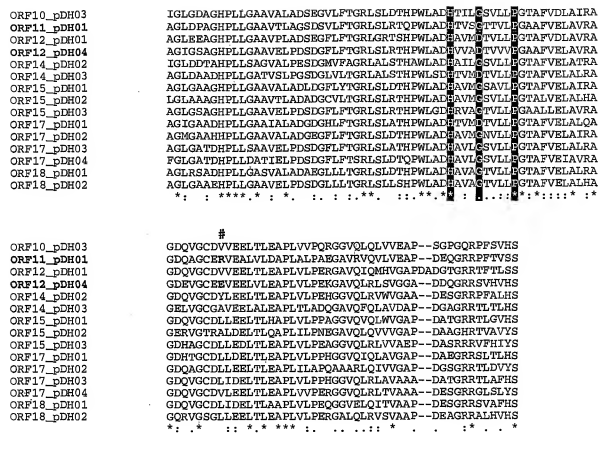

ORF10_pDH03 IGLGDAGHPLLGAAVALADSEGVLFTGRLSLDTHPWLADTTLESVLLEGTAFVDLAIRA
ORF11_pDH01 AGLDPAGHPLLGAAVTLAGSDSVLFTGRLSLRTQPWLADTTVSETTVLEGAAFVELAVRA
ORF12_pDH01 AGLEEAGHPLLGAAVPLADSEGFLFTGRLGRTSHPWLADTAVMETVLLEGTAFVDLAVRA
ORF12_pDH04 AGIGSAGHPLLGAAVELPDSDGFLFTGRLSLRTHPWLADTVVAETVVVEGAAFVELAVRA
ORF14_pDH02 IGLDDTAHPLLSAGVALPESDGMVFAGRLALSTHAWLADTAILESVLLEGTAFVELATRA
ORF14_pDH03 AGLDAADHPLLGATVSLPGSDGLVLTGRLALSTHPWLSDTTVMETVLLEGTAFVELALRA
ORF15_pDH01 AGLGAAGHPLLGAAVALADLDGFLYTGRLSLDTHPWLADTAVMESAVLEGTAFVELATRA
ORF15_pDH02 LGLAAAGHPLLGAAVTLADADGCVLTGRLSLRTHPWLADTAVMESVLLEGTALVELALHA
ORF15_pDH03 AGLGSAGHPLLGAAVELPDSDGFLFTGRLSLRTHPWLGDTRVAETVLLEGAALLELAVRA
ORF17_pDH01 AGIGAADHPLLGAAIALADGDGHLFTGRLSLATHPWLADTTVMETVLLEGTAFVELALQA
ORF17_pDH02 AGMGAAHHPLLGAAVALADGEGFLFTGRLSLDTHPWLADTAVMENVLLEGTAFVELATRA
ORF17_pDH03 AGLGATDHPLLSAAVELPDSDGFLFTGRLSLATHPWLADTAVLESVLLEGTAFVELALRA
ORF17_pDH04 FGLGATDHPLLDATIELPDSDGFLFTGRLSLDTQPWLADTAVLESVLLEGTAFVELAVRA
ORF18_pDH01 AGLRSADHPLLGASVALADAEGLLLTGRLSLDTHPWLADTAVAETVLLEGTAFVELALRA
ORF18_pDH02 AGLGAAEHPLLGAAVELPDSDGLLLTGRLSLLSHPWLADTAVAETVLLEGTAFVELALHA
*: : ****.* : *. :. : :.**. ::.**.*M : M..:.M*:*:::* :*
#
ORF10_pDH03 GDQVGCDVVEELTLEAPLVVPQRGGVQLQLVVEAP--SGPGQRPFSVHS
ORF11_pDH01 GDQAGCERVEALVLDAPLALPAEGAVRVQVLVEAP--DEQGRRPFTVSS
ORF12_pDH01 GDQVGCDVVEELTLEAPLVLPERGAVQLQMHVGAPLADGTGRRTFTLSS
ORF12_pDH04 GDEVGCEEVEELVLEAPLVLPEKGAVQLRLSVGGA--DDQGRRSVHVHS
ORF14_pDH02 GDQVGCDYLEELTLEAPLVLPEHGGVQLRVWVGAA--DESGRRPFALHS
ORF14_pDH03 GELVGCGAVEELALEAPLTLADQGAVQFQLAVDAP--DGAGRRTLTLHS
ORF15_pDH01 GDQVGCDLLEELTLHAPLVLPPAGGVQVQLWVGAP--DATGRRTLGVHS
ORF15_pDH02 GERVGTRALDELTLQAPLILPNEGAVQLQVVVGAP--DAAGHRTVAVYS
ORF15_pDH03 GDHAGCDLLEDLTLEAPLVLPEAGGVQLRLVVAEP--DASRRRVFHIYS
ORF17_pDH01 GDHTGCDLLDELTLEAPLVLPPHGGVQIQLAVGAP--DAEGRRSLTLHS
ORF17_pDH02 GDQAGCDLLEELTLEAPLILAPQAARLQIVVGAP--DGSGRRTLDVYS
ORF17_pDH03 GDQVGCDLIDELTLEAPLVLPPHGGVQLRLAVAAA--DATGRRTLAFHS
ORF17_pDH04 GDQVGCDVLEELTLEAPLVVPERGGVQLRLTVAAA--DEAGRRGLSLYS
ORF18_pDH01 GDQVGCDLIDELTLAAPLVLPEQGGVELQITVAAP--DESGRRSVAFHS
ORF18_pDH02 GQRVGSGLLEELTLEAPLVLPERGALQLRVSVAAP--DEAGRRALHVHS
: . :: *.* *** :. .. ..:: * . . :* . . *

Figure 5

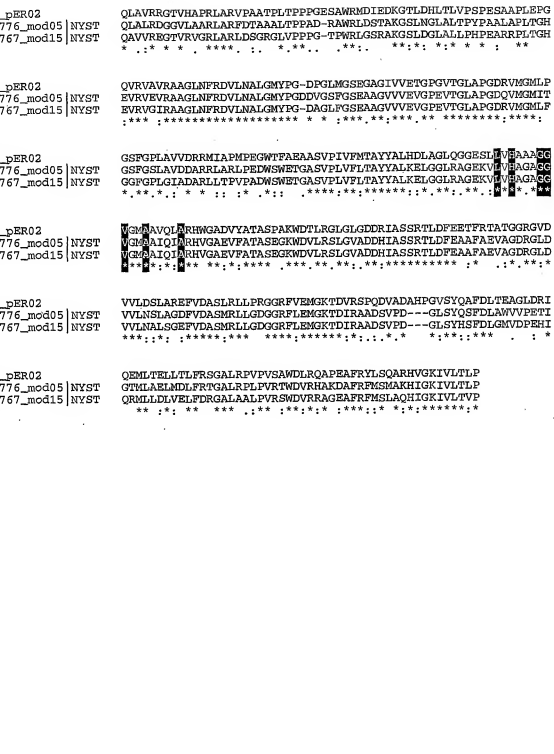
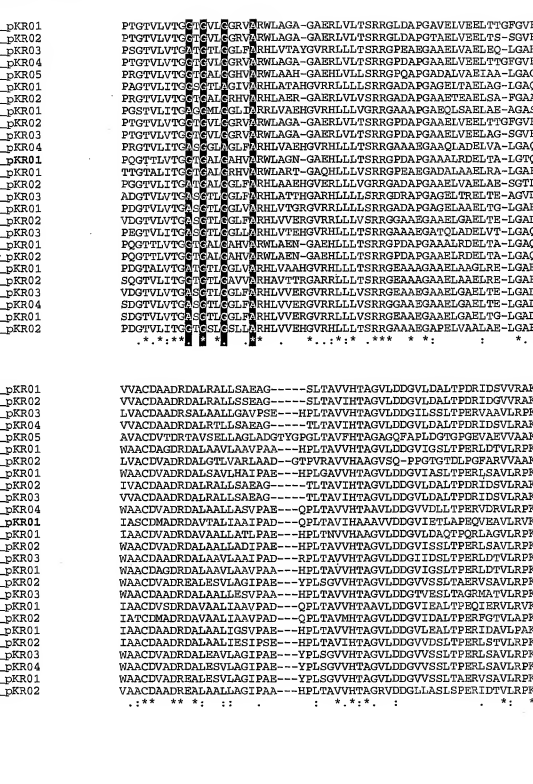


Figure 6a


ORF10_pKR01 PTGTVLVTGGTGVLGGRVARWLAGA-GAERLVLTSRRGLDAPGAVELVEELTTGFGVEVS
ORF10_pKR02 PTGTVLVTGGTGVLGGRVARWLAGA-GAERLVLTSRRGLDAPGTAELVEELTS-SGVEVS
ORF10_pKR03 PSGTVLVTGATGTLGGLFARHLVTAYGVRRLLLTSRRGPEAGGAAELVAELEQ-LGAHVE
ORF10_pKR04 PTGTVLVTGATGVLGGRVARWLAGA-GAERLVLTSRRGPDAPGAAELVEELTTGFGVEVS
ORF10_pKR05 PRGTVLVTGTTGALGGHVARWLAAH-GAEHLVLLSRRGPQAPGAADAVAELAA-LGAGAT
ORF11_pKR01 PAGTVLITGGGGTLGGIVARHLATAHGVRRLLLLSRRGADAPGAGELTAELAG-LGAQVS
ORF11_pKR02 PRGTVLVTGGTGALGRHVARHLAER-GAERLVLVSRRGADAPGAAETEAELSA-PGAAVT
ORF12_pKR01 PGSTVLITGAGGMLGGLFARRLVAEHGVRHLLLVGRRGAAAPGAEQLSAELAE-AGASVT
ORF12_pKR02 PTGTVLVTGGTGVLGGRVARWLAGA-GAERLVLTSRRGPDAPGAAELVEELTTGFGVEVS
ORF12_pKR03 PTGTVLVTGATGVLGGRVARWLAGA-GAERLVLTSRRGPDAPGAAELVEELAG-SGVEVS
ORF12_pKR04 PRGTVLITGAGGGLAGLFARHLVAEHGVRHLLLTSRRGAAAPGAAQLADELVA-LGAQVT
ORF13_pKR01 PQGTVLVTGGTGALGAHVARWLAGN-GAEHLLLTSRRGPDAPGAAALRDELTA-LGTQVT
ORF14_pKR01 TTGTALITGGTGALGRHVARWLART-GAQHLLLVSRRGPEAEGADALAAELRA-LGAEVT
ORF14_pKR02 PGGTVLITGAGALGLGFARHLAAEHGVERLLLVGRRGADAPGAAELVAELAE-SGTLAT
ORF14_pKR03 ADGTVLVTGAGGTLGGLFARHLATPHGARHLLLLSRRGDRAPGAGELTRELTE-AGVDVT
ORF15_pKR01 PDGTVLVTGAGGTLGGLVARHLVTGRGVRRLLLLSRRGADAPGAGELAAELTG-LGAEVS
ORF15_pKR02 VDGTVLVTGAGGTLGGLFARHLVVERGVRRLLLVSRRGGAAEGAAELGAELTE-LGADVR
ORF15_pKR03 PEGTVLITGAGGTLGGLLARHLVTEHGVRHLLLTSRRGAAAEGATQLADELVT-LGAQVT
ORF16_pKR01 PQGTTLVTGGTGALGAHVARWLAEN-GAEHLLLTSRRGPDAPGAAELRDELTA-LGAQVT
ORF16_pKR02 PDGTALVTGAGGTLGGLVARHLVAAHGVRHLLLTSRRGEAAGAAELAAGLRE-LGAEVT
ORF17_pKR02 SQGTVLITGGGGTLGGAVVARHAVTTRGARRLLLTSRRGEAAAGAAELAAELRE-LGAEVT
ORF17_pKR03 VDGTVLVTGAGGTLGGLFARHLVVERGVRRLLLVSRRGEAAEGAAELGAELTE-LGADVR
ORF17_pKR04 SDGTVLVTGAGGTLGGLFARHLVVERGVRRLLLVSRRGGAAEGAAELGAELTE-LGADVR
ORF18_pKR01 SDGTVLVTGAGGTLGGLFARHLVVERGVRRLLLVSRRGEAAEGAAELGAELTG-LGADVR
ORF18_pKR02 PDGTVLITGGTGSLGSLLARHLVVEHGVRHLLLTSRRGAAAEGAPELVAALAE-LGAEAT

ORF10_pKR01 VVACDAADRDALRALLSAEAG-----SLTAVVHTAGVLDDGVLDALTPDRIDSVVRAKAV
ORF10_pKR02 VVACDAADRDALRALLSSEAG-----SLTAVIHTAGVLDDGVLDALTPDRIDGVVRAKAV
ORF10_pKR03 LVACDAADRSALAALLGAVPSE---HPLTAVVHTAGVLDDGILSSLTPERVAAVLRPKVD
ORF10_pKR04 VVACDAADRDALRTLLSAEAG-----TLTAVIHTAGVLDDGVLDALTPDRIDSVLRAKAV
ORF10_pKR05 AVACDVTDRDAVSELLAGIRDGTYGPGLTAVFHTAGAGQPALDGTGPGEVAEVVAAKVA
ORF11_pKR01 WAACDAGDRDALAAVLAAVPAA---HPLTAVVHTAGVLDDGVIGSLTPERLDTVLRPKAD
ORF12_pKR01 LVACDVADRDALGTLVARLAAD---GTPVRAVVHAAGVSQ-PPGTGTDLPGFARVVVAKTA
ORF12_pKR01 WAACDVADRDALSAVLHAIPAE---HPLGAVVHTAGVLDDGVIASLTPERLSAVLRPKVD
ORF12_pKR02 IVACDAADRDALRALLSAEAG-----TLTAVIHTAGVLDDGVLDALTPDRIDSVLRAKAV
ORF12_pKR03 VVACDAADRDALRALLSAEAG-----TLTAVIHTAGVLDDGVLDALTPDRIDSVLRAKAV
ORF12_pKR04 WAACDVADRDALAALLASVPAE---QPLTAVVHTAAVLDDGVVDLLTPERVDRVLRPKAE
ORF13_pKR01 IAACDVADRDALRALLGSVPAE---QPLTAVVHAAAVLDDGVIETLAPEQVEAVLRVKVD
ORF14_pKR01 VAACDVADRDAVAALLATLPAE---HPLTNVVHAAGVLDDGVLDAQTPQRLAGVLRPKAH
ORF14_pKR02 WAACDVADRDALAALLADIPAE---HPLTAVVHTAGVLDDGVISSLTPERLSAVLRPKVD
ORF14_pKR03 WAACDAADRDALAAVLAAIPAD---RPLTAVVHTAGVLDDGIIDSLTPERLDTVLRPKVD
ORF15_pKR01 WAACDAGDRDALAAVLAAVPAA---HPLTAVVHTAGVLDDGVIGSLTPERLDTVLRPKAD
ORF15_pKR02 WAACDVADREALESVLAGIPAE---YPLSGVVHTAGVLDDGVVSSLTAERVSAVLRPKVD
ORF15_pKR03 WAACDAADRDALAALLESVPAA---HPLTAVVHTAGVLDDGTVESLTAGRMATVLRPKVD
ORF16_pKR01 IAACDVSDRDAVTALLAAVPAD---QPLTAVVHTAAVLDDGVIEALTPEQIERVLRPKVD
ORF16_pKR02 IATCDMADRDAVAALIAAVPAD---QPLTAVMHTAGVLDDGVIDALTPERFGTVLAPKAD
ORF17_pKR01 IAACDAADRDALAALIGSVPAE---HPLTAVVHTAGVLDDGVLEALTPERIDAVLPAKVD
ORF17_pKR02 IAACDAADRDALAALIESIPSE---HPLTAVIHTAGVLDDGVVDSLTPERLSTVLRPKVD
ORF17_pKR03 WAACDVADRDALEAVLAGIPAE---YPLSGVVHTAGVLDDGVVSSLTPERLSAVLRPKVD
ORF17_pKR04 WAACDVADRDALESVLAGIPAE---YPLSGVVHTAGVLDDGVVSSLTPERLSAVLRPKVD
ORF18_pKR01 WAACDVADREALESVLAGIPAE---YPLSGVVHTAGVLDDGVVSSLTAERVSAVLRPKVD
ORF18_pKR02 VAACDAADREALDALLAGIPAA---HPLTAVVHTAGRVDDGLLASLSPERIDTVLRPKAD

Figure 6b

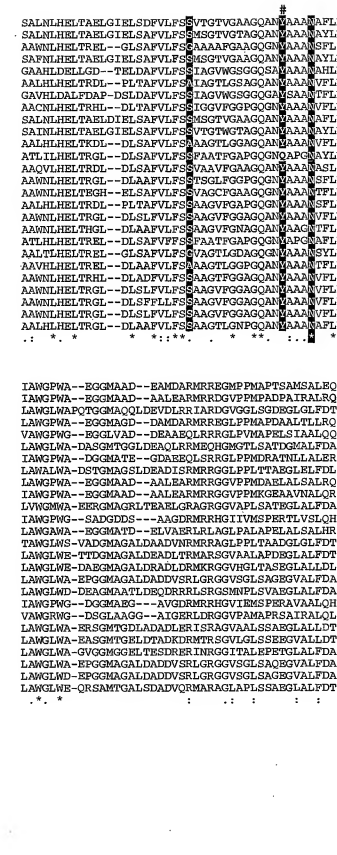

ORF10_pKR01
ORF10_pKR02
ORF10_pKR03
ORF10_pKR04
ORF10_pKR05
ORF11_pKR01
ORF11_pKR02
ORF12_pKR01
ORF12_pKR02
ORF12_pKR03
ORF12_pKR04
ORF13_pKR01
ORF14_pKR01
ORF14_pKR02
ORF14_pKR03
ORF15_pKR01
ORF15_pKR02
ORF15_pKR03
ORF16_pKR01
ORF16_pKR02
ORF17_pKR01
ORF17_pKR02
ORF17_pKR03
ORF17_pKR04
ORF18_pKR01
ORF18_pKR02

Figure 7


ORF10_pAC00 SAALRDAAPDTLDPHRPFLDLGFDSLAAVDLHARLVAGTGLRLPVTLAFDHPTPAHLARHLH
ORF10_pAC01 AAVLGHDGSDAVGAERAFKELGFDSLTSVELRNRLGAATDLRLPTTLVYDYPTSAALAEYLR
ORF10_pAC02 AAVLGHAGVENVGAGRAFKELGFDSLMAVELRNRIGSATELRLPATLIYDHPTSAALAEFLR
ORF10_pAC03 AEVLGHTDARAVDADRAFKELGFDSLTAVELRNVLKAATGLRLSPTLVFDYPTPVALARHLL
ORF10_pAC04 AAVLGHGGSEAVGAERAFKELGFDSLTAVELRNRLGAATGVRLPATLIFDYPTATALAAYLR
ORF10_pAC05 AVVLGHGATAVEAARAFKELGFDSLTAVELRNRLSTATGLRLPASLVFDYPTPAALAAHIR
ORF11_pAC01 ASVLGHASAEQVDPARAFKDLGFDSLTAVELRNRLGAATGLRLPTTLVFDHPTPTALVRHLR
ORF11_pAC02 ATALGHTSADAVAAERAFKDLGFDSLTAVELRNRLGAACGLRLPSSLVFDYPNPQALTRHLL
ORF12_pAC01 AAALGYPGPSAVEPGRSFKELGFDSLTAVELRNLLGDATGRRLPATLVFDYPTATALAGYLR
ORF12_pAC02 AEVLGHSGAEDIEAGRAFREIGFDSLTAVELRNRLGAAAELRLPATLVYDYPTPAALAVHLR
ORF12_pAC03 AAVLGHAGVESIGAARAFKELGFDSLTAVELRNRLGAVTGLRLPATLIYDYPTSGALAEYLR
ORF12_pAC04 AAVLGYGSAEHIGGEQAFKELGFDSLTAVELRNRLGAAGGLRLPATLIYDYPNPAALAQHLL
ORF13_pAC01 AAVLGHADLAAVEAGRAFKELGFDSLTSVELRNRLGAVSGLKLPASLVFDHPTPAAVAAFLR
ORF14_pAC01 AAVLGYAGPESVDPGSAFRDLGFDSLTAVEIRNLLTSRTGLRLPATLIFDYPNSLSLAAFLQ
ORF14_pAC02 AAVLGHAGPAAVESGRAFKELGFDSLTAVELRNRLNAATALRLPATLIFDYPDPTVLARYLR
ORF14_pAC03 AAVLGYAGPDDVDAARGFLDLGFDSLTAVDLRNRLTASAGLRLPVTLIFDYPSPTALAAYLA
ORF15_pAC01 ADVLGHGSPDAIDPEQAFSRLGFDSLTAVELRNRLGAAIGRRLPATLIFDHPASLTLARHLS
ORF15_pAC02 AAVLGYASPEAVEKDSSFRELGFDSLTAVELRNLLGAATGLRLPATLVFDYPTSAVLADHLR
ORF15_pAC03 AQVLGHSGAAAIEPGSAFKELGFDSLTAVELRNRLGAVTGLRLPATLIFDYPTPEALSGHLR
ORF16_pAC01 AAVLGYAGPDAVEAGRAFKELGFDSLTSVELRNRLNAASGLKLPPTLVFDHPTPVLARHLR
ORF16_pAC02 ADALGHPTTDEVGAGRAFKELGFDSLTALELRNRLNAATGLRLPATLVFDHPTPTILAEFLR
ORF17_pAC01 AAVLGYPGPEAVDPGRAFKELGFDSLTAVELRNRLGSATGVRLPATLVFDYPTPNALSAFLR
ORF17_pAC02 AAVLGHATPDAVEPTRAFKDLGFDSLTAVEFRNRLGATAGIRLPATLVFDYPTPTVLAGYLK
ORF17_pAC03 AAVLGHASTDEVPADRAFKELGFDSLTSVELRNRLGATTGERLSATLVFDYPTPHALAEFLR
ORF17_pAC04 AAVLGFAGPEAVDPARSFSEVGFDSLTAVELRNRLGAATGVRLPATLVFDYPTPDALVEYLR
ORF18_pAC01 AAVLGLAGPEAVDPARSFSEVGFDSLTAVELRNRLGAATGVRLPATLVFDYPTSLALADFLG
ORF18_pAC02 AAVLAYPSPDAVGESQEFLELGLDSLTAVELRNQLNAATGLRLPATLFDHPTPALVAERLR

Figure 8

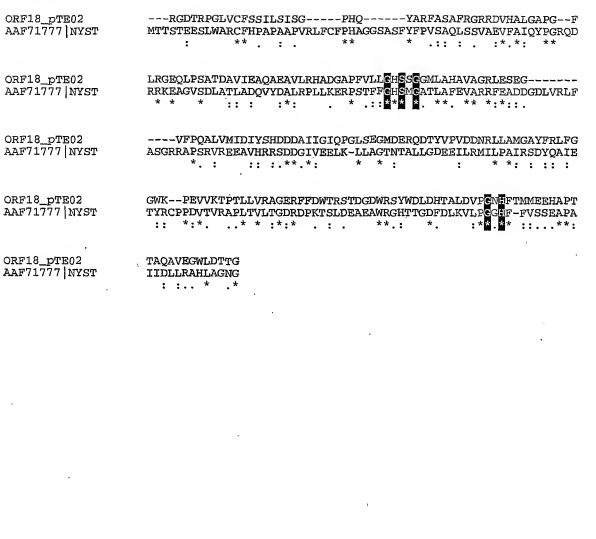

ORF18_pTE02
AAF71777|NYST
---RGDTRPGLVCFSSILSISG-----PHQ------YARFASAFRGRRDVHALGAPG--F
MTTSTERSLWARCFHPAPAAPVRLPCFPHAGGSASFYFPVSAQLSSVAEVFAIQYPGRQD
ORF18_pTE02
AAF71777|NYST
LRGEQLPSATDAVIEAQAEAVLRHADGAPFVLLGHSSGGMLAHAVAGRLESEG-------
RRKEAGVSDLATLADQVYDALRPLLKERPSTFFGHSMGATLAFEVARRFEADDGDLVRLF
ORF18_pTE02
AAF71777|NYST
----VFPQALVMIDIYSHDDDAIIGIQPGLSEGMDERQDTYVPVDDNRLLAMGAYFRLFG
ASGRRAPSRVREEAVHRRSDDGIVEELK-LLAGTNTALLGDEILRMILPAIRSDYQAIE
ORF18_pTE02
AAF71777|NYST
GWK--PEVVKTPTLLVRAGERFFDWTRSTDGDWRSYWDLDHTALDVFGNHFTMMEEHAPT
TYRCPPDVTVRAPLTVLTGDRDPKTSLDEAEAWRGHTTGDFDLKVLPGGHFFVSSEAPA
ORF18_pTE02
AAF71777|NYST
TAQAVEGWLDTTG
IIDLLRAHLAGNG

Figure 9

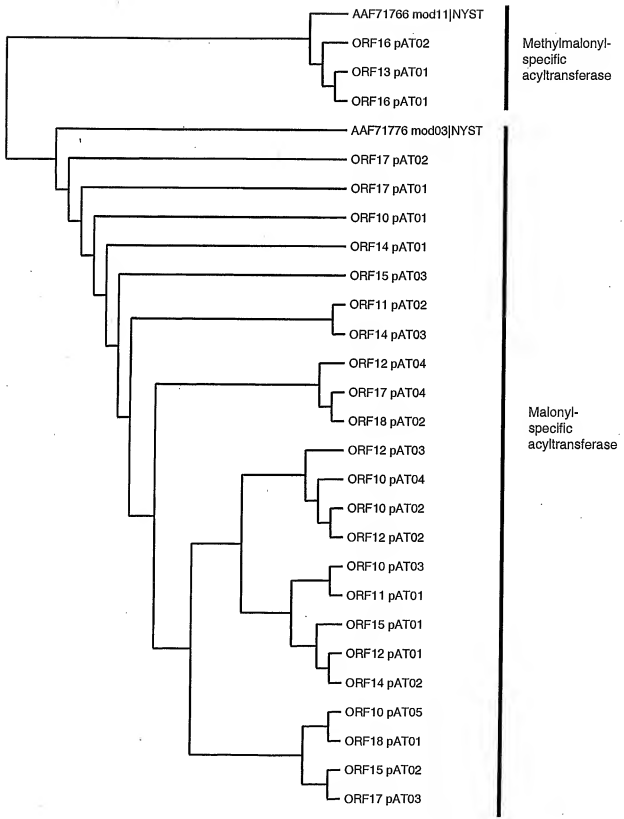
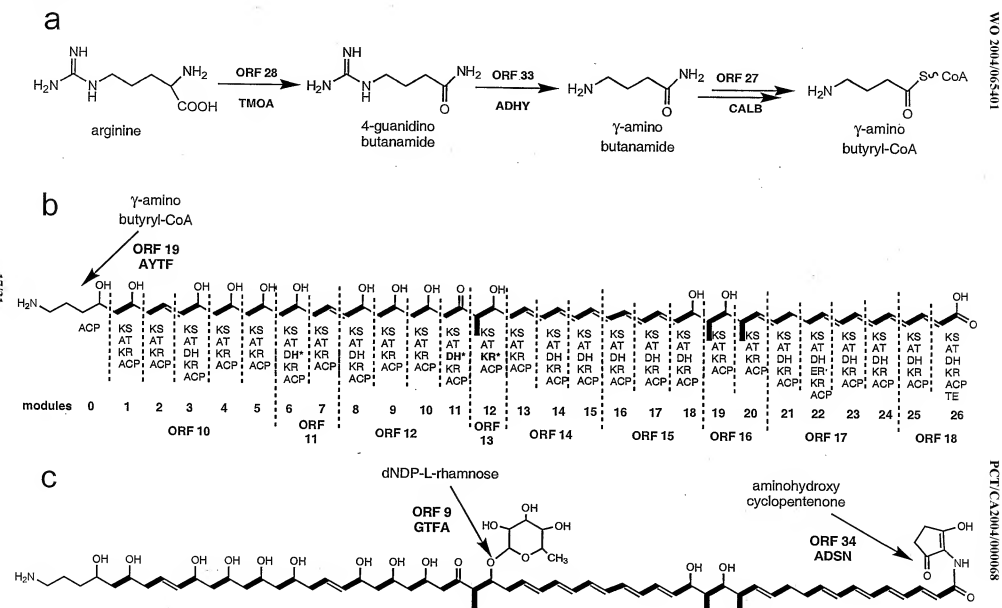

AAF71766 mod11|NYST
ORF16 pAT02
ORF13 pAT01
ORF16 pAT01
Methylmalonyl-specific acyltransferase
AAF71776 mod03|NYST
ORF17 pAT02
ORF17 pAT01
ORF10 pAT01
ORF14 pAT01
ORF15 pAT03
ORF11 pAT02
ORF14 pAT03
ORF12 pAT04
ORF17 pAT04
ORF18 pAT02
ORF12 pAT03
ORF10 pAT04
ORF10 pAT02
ORF12 pAT02
ORF10 pAT03
ORF11 pAT01
ORF15 pAT01
ORF12 pAT01
ORF14 pAT02
ORF10 pAT05
ORF18 pAT01
ORF15 pAT02
ORF17 pAT03
Malonyl-specific acyltransferase

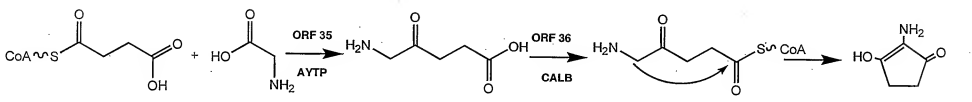
Figure 10



WO 2004/065401
PCT/CA2004/000068

Figure 11

a



succinyl-CoA glycine 5-aminolevulinate 5-aminolevulinate-CoA aminohydroxy cyclopentenone

b

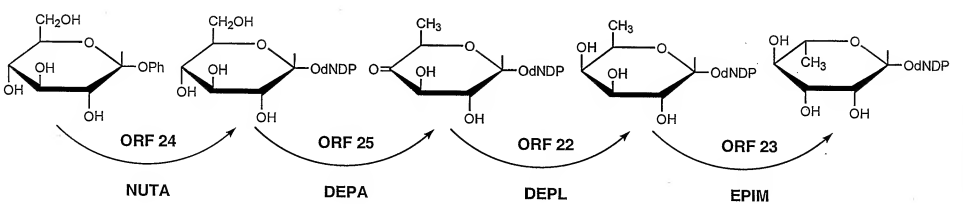


Figure 12a

Figure 12b

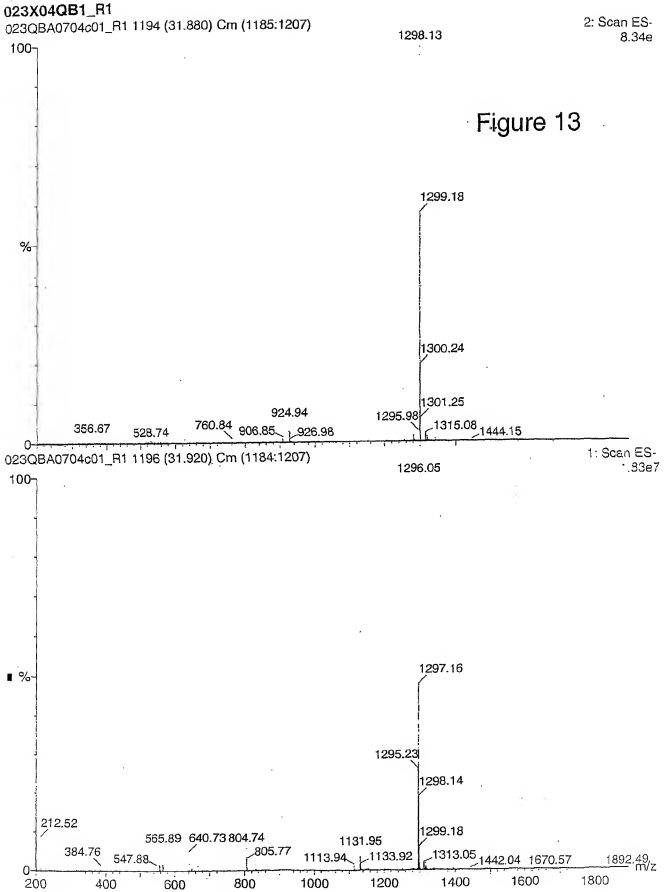
Figure 12c

Figure 12d

Figure 12e

Figure 12f

Figure 13


023X04QB1_R1
023QBA0704c01_R1 1194 (31.880) Cm (1185:1207)
1298.13
2: Scan ES-
8.34e
100
%
1299.18
1300.24
1301.25
924.94
1295.98
356.67
528.74
760.84
906.85
926.98
1315.08
1444.15
0
023QBA0704c01_R1 1196 (31.920) Cm (1184:1207)
1296.05
1: Scan ES-
.33e7
100
%
1297.16
1295.23
1298.14
212.52
565.89
640.73
804.74
1131.95
1299.18
384.76
547.88
805.77
1113.94
1133.92
1313.05
1442.04
1670.57
1892.49
m/z
0
200
400
600
800
1000
1200
1400
1600
1800

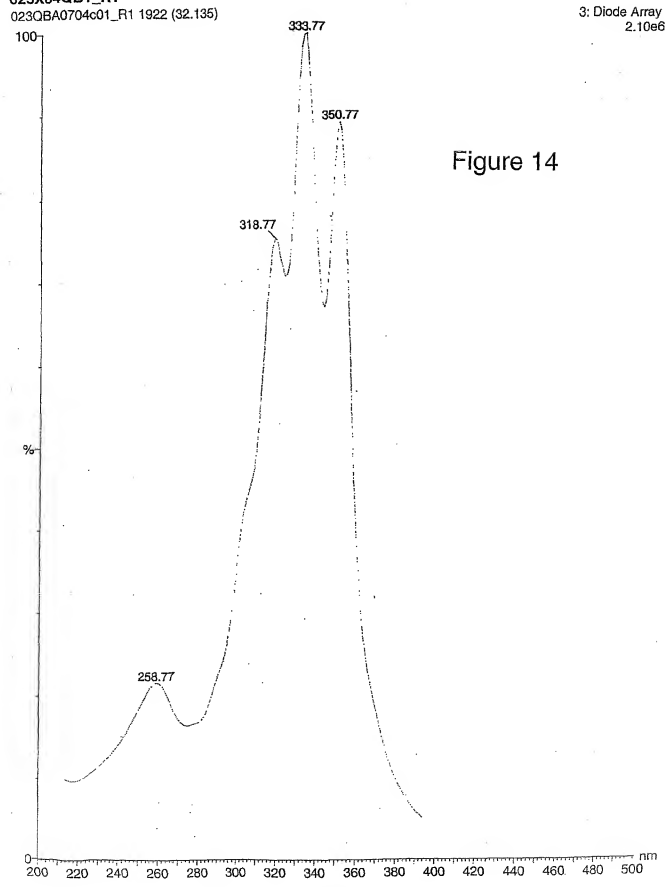

023X04QB1_R1
023QBA0704c01_R1 1922 (32.135)
3: Diode Array
2.10e6
333.77
350.77
318.77
258.77
100
%
0
200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500
nm

Figure 14

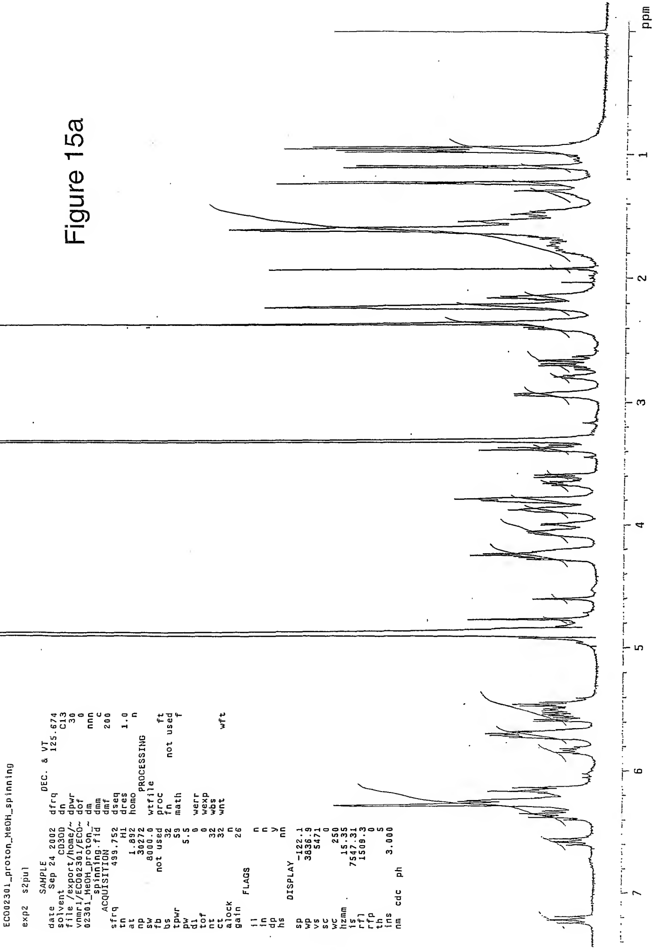


Figure 15a

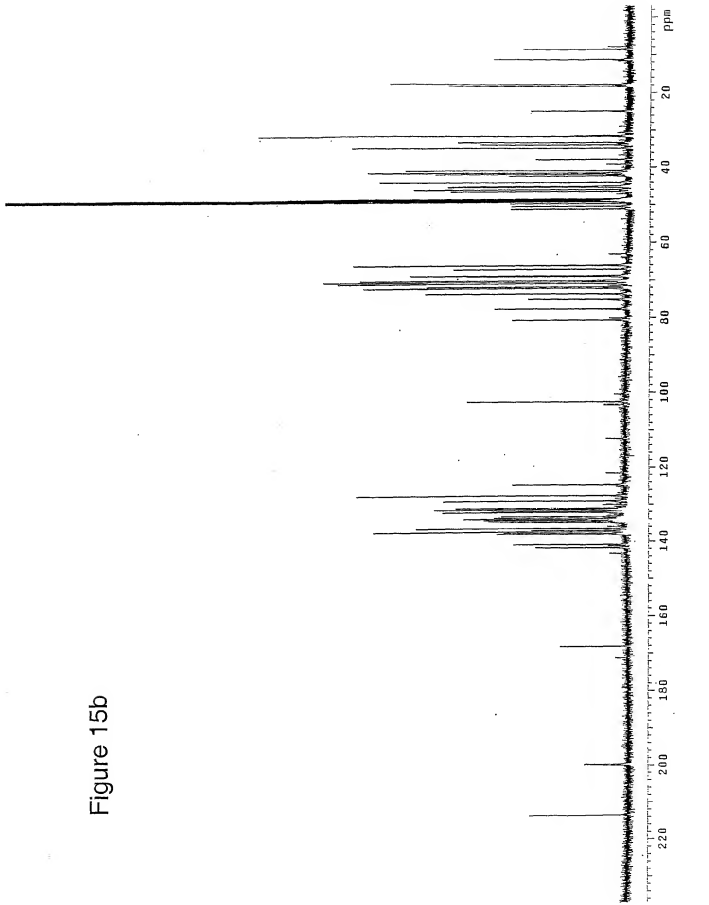


Figure 15b

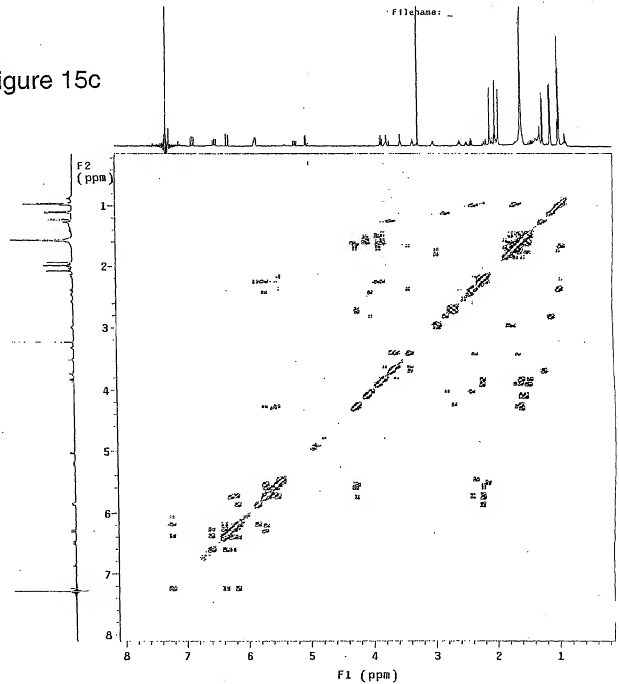
STANDARD PROTON PARAMETERS

Pulse Sequence: gDQCOSY

Solvent: CD3OD
Ambient temperature
File: ECO_02301_gDQCOSY_CD3OD_high_res
INOVA-500 "resonance"

Relax. delay 1.000 sec
Acq. time 0.128 sec
Width 3988.0 Hz
2D Width 3988.0 Hz
32 repetitions
2 x 500 increments
OBSERVE H1, 499.7495714 MHz
DATA PROCESSING
Sq. sine bell 0.096 sec
Shifted by -0.064 sec
F1 DATA PROCESSING
Sq. sine bell 0.037 sec
Shifted by -0.025 sec
FT size 8192 x 8192
Total time 10 hr, 48 min, 6 sec

Figure 15c



STANDARD PROTON PARAMETERS

Pulse Sequence: gHSQC

Solvent: CD3OD
Ambient temperature
File: ECO-02301_CD3OD_gHSQC_12_02
INOVA-500 "resonance"

Relax. delay 1.000 sec
Acq. time 0.128 sec
Width 4001.6 Hz
20 Width 21361.8 Hz
64 repetitions
2 x 128 increments
OBSERVE H1, 499.7495714 MHz
DECOUPLE C13, 125.6727610 MHz
Power 40 dB
on during acquisition
off during delay
WALTZ16 modulated
DATA PROCESSING
Gauss apodization 0.059 sec
F1 DATA PROCESSING
Gauss apodization 0.022 sec
FT size 2048 x 2048
Total time 5 hr, 28 min, 9 sec

Filename: _

Figure 15d

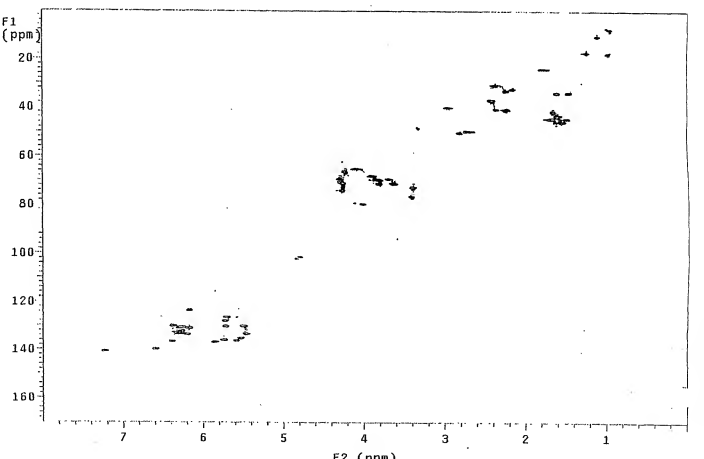
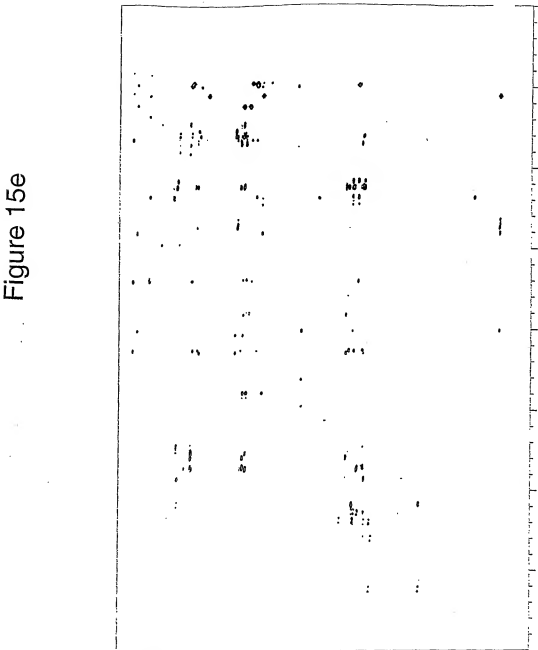


Figure 15e


STANDARD PROTON PARAMETERS
Pulse Sequence: gHMBC
Solvent: CD3OD
Ambient temperature
File /export/home/chem/gHMBC_12_02
INOVA-500 "rresonance"
Relax. delay 1.000 sec
Acq. time 1.000 sec
Width 4021.6 Hz
20 Width 24891.7 Hz
1H nucleus 500 MHz
400 increments
DATA PROCESSING
FT size 2048 x 4096
Total time 17 Hr, 18 min, 49 sec
Filename: _
F1 (ppm)
F2 (ppm)

STANDARD PROTON PARAMETERS

Pulse Sequence: TOCSY

Solvent: CD3OD
Ambient temperature
File: 02301_CD3OD_TOCSY_mix_08
INOVA-500 "resonance"

Relax. delay 1.000 sec
Mixing 0.080 sec
Acq. time 0.128 sec
Width 4001.6 Hz
2D Width 4001.6 Hz
32 repetitions
2 x 200 increments
OBSERVE H1, 499.7495714 MHz
DATA PROCESSING
Gauss apodization 0.059 sec
F1 DATA PROCESSING
Gauss apodization 0.046 sec
FT size 2048 x 2048
Total time 4 hr, 29 min, 30 sec

Filename: _

Figure 15f

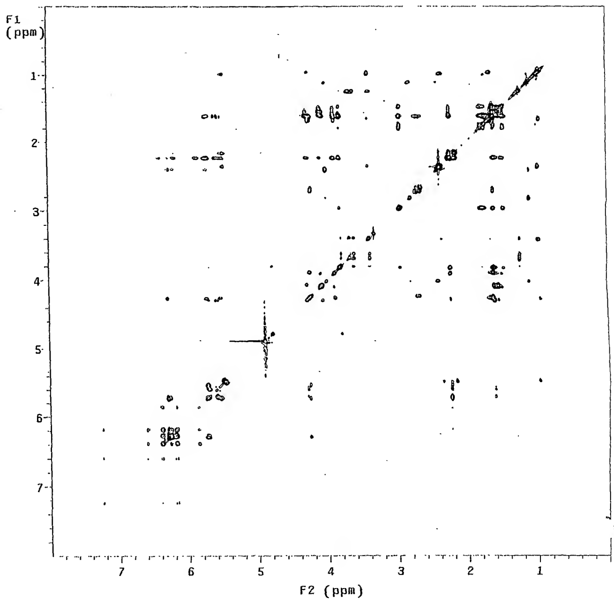


Figure 16

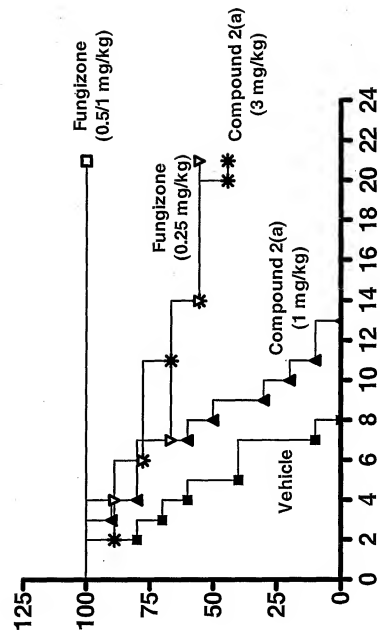

Effect of Compound 2(a)
on systemic Candidiasis in vivo
Percent survival
125
100
75
50
25
0
0 2 4 6 8 10 12 14 16 18 20 22 24
Days post-inoculation
Fungizone
(0.5/1 mg/kg)
Fungizone
(0.25 mg/kg)
Compound 2(a)
(3 mg/kg)
Compound 2(a)
(1 mg/kg)
Vehicle

Figure 17a

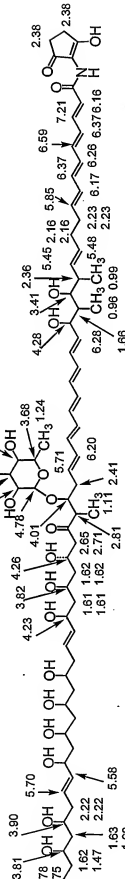


Figure 17b

SEQUENCE LISTING

<110> Ecopia BioSciences Inc
Bachmann, Brian O.
McAlpine, James B.
Zazopoulos, Emmanuel
Farnet, Chris M.

<120> POLYENE POLYKETIDES, PROCESSES FOR THEIR PRODUCTION AND THEIR USE AS A PHARMACEUTICAL

<130> 3004-8PCT

<150> USSN 60/441,123
<151> 2003-01-21

<150> USSN 60/469,810
<151> 2003-05-13

<150> USSN 60/491,516
<151> 2003-08-01

<150> USSN 60/494,568
<151> 2003-08-13

<160> 78

<170> PatentIn version 3.0

<210> 1
<211> 11740
<212> DNA
<213> Streptomyces aizunensis

<400> 1
gatcatggcc ggcgaggtgg tcgcgggcgg ggcgaatccg aaggtcacgg tcctcccttc 60

gggttacgcg cgccgctgac gggcacggct gggttgcggg cgcgccgcag cgcggccctc 120

aagagtgccg acgagccgag cgggaacact ccaattctcg cgcggcccgc gaggatgcgg 180

caacgagcaa ttggcgccgc ggaccgtaat tggccggtat gccgttcata tccttgcccc 240

gttacgccgt cgatgacgca tccggtgccg cccggaccgc cggtaccagc ggaaacacct 300

cccgcgcggc ggcccgctgg agccgcggag atccaccgga caccccctgg gcctggcgga 360

gtccgtgcgt gccgcgtgga ttcgccgatt gtcggtggga tcgggttgca tgggggcatg 420

gacaacctgg agctccgtcg tgaagccgat gccatcctcg ctgagctggt cggtgcccct 480

gggggttcgg cgcggctgcg ggaggaccag tggcaggcgg tcgcggccct ggtggaggag 540

cgccggcggg ccctggtggt gcagcgcacg ggctggggca agtccgcggt ctacttcgtc 600

gccaccgctc tgctgcgccg gcgcggctcc gggccgacgg tgatcatttc tccgctgctg 660

gcgctgatgc gcaaccaggt cgaggcggcc gcgcgggccg ggatccaggc gcgcacgatc 720

aactcggcca acccggagga gtgggaaacc atctacgggg aggtcgagcg cggcgagacc 780

gatgtgctcc tcgtcagccc cgagcgcctc aactccgtgg atttccgcga ccaggtactg 840

cccaagctgg cggccacgac gggtctgctg gtggtcgacg aggcgcactg catctccgac 900

tggggccacg acttccgccc cgactaccga cggctgcgca cgatgctggc ggagctgccg 960

gagggcgtgc cggtcctggc cacgacggcg accgcgaacg cgcgggtgac cgcggacgtg 1020

gcggagcagc tgggcacgca cggcgagcac gccctggtcc tgcgcggacc gctcgaccgg 1080

gagagcctgc ggctgggagt gctgcagctg ccggacgcgg cgcaccggct ggcctggctg 1140

ggggaccggc tggcgcacct gccgggttcg gggatcatct acacgctgac cgtggcggcg 1200

gcggaggagg tcgcggcgtt cctgcggcaa cgcgggtatc cggtggcttc ctacaccggg 1260

aagacggaga acgccgaccg gttgcaggcg gaggaggatc tgctggcgaa ccggggtgaag 1320

gcactggtgg cgacctcggc gctgggcatg gggttcgaca agccggacct ggggttcgtg 1380

gtgcacatgg ggtcgccctc gtccccgatc gcctactacc agcaggtggg gcgcgcgggg 1440

cgtggggtgg atcacgcgga cgtgctgctg ctgccgggcc gggaggacga ggcgatctgg 1500

gcgtacttcg cctcggtggg cttcccgccc gaggagcagg tccggcgcac cctggacgta 1560

ctggcgcagg cgggccgccc gctgtcgctg cccgcgctgg agccgctggt ggacctccgg 1620

cgctcgcgcc tggagacgat gctgaaggtc ctggacgtgg acggcgcggt caagcgcgtg 1680

aagggcggct ggaccgccac cgggcagccg tggacgtacg acgcggagcg gtacgcctgg 1740

gtcgcgaagc agcgggcggc ggagcagcag gccatgcggg actacgtggc gaccacgggc 1800

tgccggatgg agttcctgca gcggcagctg gacgacgaga aggcggtccc gtgcggccgc 1860

tgcgacaact gcgccggatc ctggctggag gcggtcgtgt cgcccgcggc cctcgcggcc 1920

gcggcgggcg agctggaccg cgcgggggtc gaggtcgagt cccgcaagat gtggccgacc 1980

gggctcgccg cggtcggcat ggacctgaag ggccggatcc ccgcgggcca gcaggccgtc 2040

accgggcgcg cgctcggcag gctgtcggac atcggctggg gcaaccggct gcgccccctg 2100

ctgtcggcgc aggccgcgga cgggccggtt ccggacgatg tgctggccgc cgtcgtgacg 2160

gtgctcgccg actgggcccg ctcgccgggc ggctgggcga gcggcgggcc ggacgcgatg 2220

gcgcggccgg tggggatcgt cgccatgccc tcccgtaccc gcccgcggct ggtcgcctcg 2280

ctggccgagg gcgtggcccg ggtcggcagg ctcccgctgc tgggcagcct cgcctacacc 2340

ccgcaggccg acgtgtacgg ggcgcaccgc agcaactcag cccagcggct gcgcgccctg 2400

gccgactcgt tcaccgtgcc cgaggaactc gccgcggccc tggccgccgc tcccggcccg 2460

gtcctgctcg tcgacgacta caccgactcc ggctggaccc tggccgtggg cgcacgcctg 2520

ctgcgccagt ccggcgcggg cggcgtgctc ccgctcgtcc tcgcgctggc cgggtaggcg 2580

gactccaccg gcctcggcct atcgccaacc gacgggggcc ggcaagatca aaacaaccgc 2640

ccgtaaagca aacgtaaaga tgtggcttct ttgggaagtc gcgtatgggc ctgttttgag 2700

ccacgcggcg gaagtcaccc ctggcgggat ccgtggtggc gcattcggtg cggacggccg 2760

aacgggccgt cgtcgctccc gttcgggccg gggggccctg tcgtcgcacg gggagagcga 2820

atgccggccg gggctgcgga ccgggaggtt ccagccaggg taggggtaga aagtaggggt 2880

actccccgcc ttgatcgtcc tggtagacat gacacatccg aaacgcgcgt gcggaagtgg 2940

cggaagggtt cgacccgtcg aacgggcgcg ctgcatctgg ggcttgaaca gggagtttca 3000

gtccgttgaa taagcaagaa actagcctct gggttcgccg ctaccacgct tcggacgaaa 3060

gccggatcca attggtctgt ctgccgcacg ccggtggctc ggcctccttc tacttcccca 3120

tgtcccagtc gctggctccg gcgatggacg tcctctcggt ccagtacccc ggcaggcagg 3180

accgcaggga cgagcccggg atcgtggaca tcggcgccta cgcggacgcc ctgaccgagc 3240

aactcgtacc gtggctcgac cggcccctgg ccttcttcgg ccacagcatg ggtgcgatcc 3300

tcgccttcga ggtgacgcgc aggctggagc gtgaccacgg cgtcactccg gagcacatct 3360

tcgcttccgg ccggcgctcg cccgccagtt tccggcacga gaccgtgcac ctgcgggacg 3420

acgacggaat cgtggcggaa atgcgggaac tcagcggaac cgacgcgaag atactcggca 3480

acgaggaaat cctccgcatg gtgctccccg cgattcgaag cgactacacc gccatcgaga 3540

actaccgtgc cgcgccggaa gacgtcgtgc gtactcccat cacggtgctg accggtgacg 3600

cggacccgag gaccagccgg gaagaggcgg acgcctggaa ggcgcacacg accggcggat 3660

tcgatctgca ttccttcccc ggtggacatt tcttcctggc gaatcaccag gagaagatca 3720

tgggaattat ttcggaggaa ctctccgcgc cggctcgcat ggcgtgagca gagagctgtg 3780

gaccaggccg gggaaacccg gctcgcccct tgccgacctc caccgcgatg gcggagccga 3840

gaagccgaat gaccaacggc cgcggtggcg atcgaaaggg gcaggccgcg gtgacggccc 3900

gccggtgcac accgtgcacc ggcacaccaa gcggtgcggc ggcggcttcg ccgggcgccc 3960

accgggcccg ttgcgaagtc ttcgcaagtc gtgcagttcg ggggaaagga agcccgtggc 4020

ggttaggctc gtcgagcgcg agaagcagct ggaaacgctg aaggaactac tcggcagcgc 4080

agtccgtggc cgagggcggg tcgccgtcat cagcggggca gtcgccggcg ggaaaacgag 4140

tctgctggaa atcttcaccg aagaggcgat ctccgcgggc gcgctggtgc tggaagccac 4200

gggctcccgg gcggagcgct atctgccctt cggaattctg cgcagaatcc tcgacagcgc 4260

ggcgcccctg tcgcccgaga tccacgccta cgccaccgag ctgctggacc gcgtcagcgc 4320

cgggacgacg gacgccgaag gcgccgtcga ggccggtatg cgcgtcctgc cccatgtcgc 4380

caccgcactg ttaaggatcg cccggaaccg gaccgtcgtc atagccatcg acgacgtcca 4440

ccacggggac gaactctccc tcgccttcct gctgtgcctc gcccgccgag tgcgccaggc 4500

gggcgtcctg atcgtgctca ccgaagccgt ccggctgcgg tccgcgcaac tcgccttcca 4560

cgccgaactg cagcgccagc ccaactgcac cagcctccgg ctgcccctgc tcaccacgcg 4620

cggcaccacc cgcgtcctcg ccgagcactt ctccccctcg acggcgcaac ggctgtccgc 4680

cgagtgccag gagaccaccg gcggcaatcc actgctggtc agggcgctga tcgacgacgg 4740

cctcacggcg ctcggagaca gcgagccctt ccagcggctc gcccccgccg aaaccttcga 4800

acgcgccgtg ctcgactgcc tgcaccgcgg cgaccccgag ctgctgaccg tcgcccgggg 4860

cgtcgccgta ctcggtagcg cctgctcctt ggccctgctc aacgggatcg tcgacctgca 4920

cgccaaggcc accgaacagg cccttcagga cctcagccgg tgcgccgtcc tgcaccacgg 4980

ctccttccgc gacccggcgg cccgtaccgc cgtcctggaa gccactccgc ccgcggcgct 5040

gtccgccctg cacctgcgca ccgcgcgact cctgcaccag gaaggcgcga cggcgctcga 5100

tgtcgcccgc cacctcctcg ccgcccgcaa gaacgtcgag gactgggcga tccccgtcct 5160

ccaggaggcg gtcgagtacg ccctcgtcga ggacgagcac gaactcgccc tgcggtgcgg 5220

ggaactggcg gtcgcctcct gcgcggaggg cccccgacac gccgccctga agtcccgcct 5280

ggcgagcatc gtctggcgca gcagcccggc cgccgctgaa gggcatctgc ggcagctgtc 5340

ccgcgaactc gccgccggcc ggctcgccga ccgcgatctc gtccaggccg tgtcgctcct 5400

ggcgtggatg ggggagtccc gggggccggg cgaggcggta ctgcgactgc agcggaccga 5460

cagcgaggcc gaggcggccg gacgggcgcc cgcctacgac ccgggcacgc tcaccgccgc 5520

acagagctgg ctctcgatgg tcagcccgcc ggccgcgcac ctcttcgacg ccgtggaacc 5580

gcgccggaca acgctgtcag gcgcgccggg ggcgctgccc ggcgcggggc ccgacaccgt 5640

cccctacgac atgcccgaca acgcctacgt ccaggccgcc gacgccgtcc gcaccgccct 5700

gcgcggcgga acccaggccg acgccgccgt cagcaaggcc acccgggtgc tccagcgcta 5760

ccacctgagc gaccgcaccc tccagccgct cgtcttcgcc ctcctcgccg tcatctacgc 5820

gggtcgcctc gacctcgcgt ccgcctggtg cgaacgactg ctcggcgagt gctccgcccg 5880

caacgccccg acctggcagg ccgccctcgg tgtggtccgg gccgagatcc tgctgcgcca 5940

gggcgatctg cccggtgcgg ccgcccaggc ccgccacgcc atgtcccgga tctccctgca 6000

gagctggggc gtgggcatcg cgctgccgct ggccgtcctc gtcgaggccg aggtccagat 6060

gggcgaccac gaggaggcga tgagcctgct cgaacagccg gtgccccagg ccatgttcga 6120

caccctggcc ggcctgcact acctcagggc ccgcggccgc tgccacctgg ccaccggccg 6180

ctaccacgcc gccgtgcggg acttcctgaa ctgcggcgag ctgatgcagg cctggggcgt 6240

ggacggggcg gagctggtgc cgtggcggct ggacgccgcc gaggcgtggc tggccctcgg 6300

caacgtcgcg cgcgccaagg agtacaccga gcagcagaag cagcgcgaga cggggcccgt 6360

gggcagccgg acgcgtggct ccctgctgct cacgctcgcc cacaccggcg gtgacctcac 6420

ggtccggctc aagcggctcg tcgaggccgt cgagaccctg gaggagggcg gggaccggct 6480

ccagctggcg gtggcgctgg gggagctggg ccgcggctac cgtgcgctgg gcgacttcaa 6540

ccgggcccgg atgctggtgc gcaaggcctg gcacgtcgcc aagtcctgcg gcgccgaacc 6600

gctgtgccag cagttcatgc cggggcaggt cgacggcgag gccggtgcgc agagcggccg 6660

ggaggcggag cttcccagcg aggtcgaggt cctgtccgag gccgaggcgc gggtcgcgct 6720

gctggcggcg cgcggccaca ccaaccgtga gatagcgacc aagctctacg tcacggtgtc 6780

cacggtcgag cagcatctga cgcgcatcta ccgcaagctg aaggtgaagc ggcgccgcga 6840

tctgcccgcc cggctgtcgg acctgagcct gccgagcatc gcctgaccgc gcccgtcgcc 6900

gggagcgcgt tgcgggagcg cgttgcccgg agcgcggcgc cacgcgcggc gcccgccgcc 6960

cgcgggccgc acccgtcagg acagcaggcc gagcttcagt gccgtgatca ccgcggccgt 7020

ccggtccgag accgacagct tcttgaacga gcgcagcaga tgcgtcttca ccgtcgcctc 7080

gctgatgaac agctggcggc cgatgtccgc gttggtcagc ccgaggctga ccaactggag 7140

cacctcgcgc tcacggtccg acagcgcggg cggctccacc acccgggccc ggaacagctt 7200

gggggcgagc gacggcgtca ggaccgtctc accgcgggcc gccgccttta ccgcctgcac 7260

cagttcgtcg cgcgagctgc ccttgagcag gtagcccgcc gcgcccgcct ccacggcccg 7320

caggatgtcc gtgtcgctct cgtacgtcgt cacgatcacc accttggtgg ccggcgcgac 7380

gcgcagcagg tggccggtgg tctccacccc gtccatcccg cccatctgaa ggtcgagcag 7440

gacgatgtcg ggagcaagtc tggtgaccat cgcgatcgcc tcctcgcccg agtcggcctg 7500

cccgacgacg ctcacgccgt cggcggattg cagcatcgag ctgagaccct cccgtacgac 7560

cgggtggtcg tcgaccagca tcacaccgat cgtcttgtca gcgctcatcg gcttcctctc 7620

ccttcgcggg cacgggcacc gtcacttcga tggtggtgcc ctgtccgggg ctgctgacca 7680

cggtcgccgc cccgctgatc tcgtgtgcgc gagtctgcat gccgcgcagc ccgcttcccc 7740

gctggtcccc ggtgacggtg aacccgggtc cgtcgtcccg tacgagcagc cgtacggtgt 7800

cctgttcgta cacgagccgg atctcggccg cgcgtgcctt ccccgcgtgc ttgcggatgt 7860

tcgcgatggc ctcctggagg gaacgcagca ggaccacgct gatcgccatc ggcagttccc 7920

gctcgtctcc ttcgacggtg acgtgcgccc gcatgccggt ctgcgccgtc aggccctcgg 7980

cctgccgccg cgtcgcctgc acgagcgagg actcctgcag cgcgggcggg gtcagctcgg 8040

tgacgaactc gcgggcttct cccaggcttt cgcgggccac gcggcccgcc agtgccagat 8100

gcgccctcgc ccggtccggg tcggccgtga agtcggtctc ggcggcctgt acgaggctga 8160

tgatgctggt gaggccctgg gcgagggtgt cgtggatctc ccgggcgagc cgctcgcgct 8220

cggcggagac ccccgccttg cgcgacagcc gggcgacttg cgcacggttg cggtgcaact 8280

cctcgatgag ctcggcccgg tcacggctct gccgggtcac ccgggtgatc cacagcccga 8340

gcatgaccga cagggcgatg ccgaggagcg aggtcggcag gacggccagg atgtcgcggc 8400

tcagggtgcc gccgcgcagc cacaccacga tgaccggaac cagattggcc agcgtgacca 8460

cggcgatggc cggcgaggtc gccaggctca tcatcagcat cgggaccacg gcgaacagcg 8520

cgaacgaggc cgcgaggtcg aagaccacgg ccaccgcgaa cagcacgaac aggccgacgg 8580

agaagacgac gctgcgccgg acgggcccct ggccctcgtg gaccatggtg ctgcgcccca 8640

gggccgcgta ccagggcacg gccgcggtca gcgcggccat ggccacggcc cggtggacct 8700

gttcaccgtc ggaggtgaac agcagcatgg tggtgacggc gtacgagacc gcgaagagcg 8760

cgtcccacag gccgaaccac cgggctcccg cctcgggcgc gtcgtcctgg ccgtctgtcg 8820

cctgcgccgc gggggattca gtgctcaccc gacaagtcct atcacttcgg tcgggcacgg 8880

tacgagggcg gcccggcgcc gtccaccgtg tccaccggtc ggtggacagc cgaacccact 8940

ggtcggttgt cctcgcgtcc cttgcccgcc gcctaacgtt gcaggtgaga ggcacgaagc 9000

gaccgcactg ccggagagaa ggcagtgccg aggaagagga agaggtcatc ccctgagccc 9060

gttcttgaac acactgatcg ccagcgggac gatcttggcc gtcattctgt cgaccgacct 9120

cggcacccgc aaagtcacca cgacgcggat gcttccttcg ctcctcgcgg tcgtcgtgat 9180

cctcgcgctc ctcgtgcaca cactgccgct cgacggcaac gacccctcgc tccaactggc 9240

gggcatcggc gccggtatca tctgcggact ggccgccacg gcgctcctcc ccgcccaccg 9300

gaacgcttcc ggtgaggtct ccaccaaggg cggtatcggt tacgcgctgg tgtggaccgc 9360

gctgtccgcc tcgcgtgtgc tcttcgccta cggttcacag cactggttca gcgagggcat 9420

cgtccggttc agcaccgact acaagctcag cggacaggcc gtctactcca acgctttcgc 9480

cttcatggcc ctggccatgg tgctgacgcg gaccgccgtc ctgttgaaca cgcgccgccg 9540

gctgcgcggc gggcagcttc ccgcggccga caacacggcc ccacatcagg cgagttccgc 9600

caatacgcac tgacatgacg gagcgtcaga tccggcttgg gtgcaagatc gtctcagaac 9660

tagggtgaag cagtgaaaca catgcatgat gtcaggctcc ggccccgcg caatcgtgtc 9720

gactcccggg cagtgggctg gtggacggtc cagtccgcga tgtacgccct gcccctgccg 9780

atcaccttcg gcgtgctgta cctgtgcatc ccgcccgcca ggccgttctt cggctgggcc 9840

ttcctgatct cgctcgtacc gggcctcgcc tacatggccg tcatgcccgc ctggcgctac 9900

cgggtgcacc gttgggagac caccgacgaa gccgtctacg cggcgtccgg ctggctctgg 9960

cagcagtggc gggtcgtgcc gatgtcccgc atccagacgg tggacaccct gcgcggaccc 10020

ctccagcagc tcttcggcct ctccggcatc accgtcacca ccgcctccta ctccggcgcc 10080

gtgaagatca agggaatcga ccaccggacc gcgcgggacg tggtcgagca cctcaccagg 10140

gtgacccagg ccacccccgg agacgcgaca tgagccacga caccggacag tgggaggcca 10200

ccgcgacctc ccacggcgcc gccgaagacc ccgagtggag caggctcagc ccccgactgc 10260

tgctggtcaa cctgagcatg ctcgccggcc cgctcgccct gttcgccgtc acggtcgccc 10320

tgaccggcgc caacctccag gccctcatct ccctcggctc cctgctgatc gtcttcctgg 10380

tcatcaccgg gatcagcacg atgcggctgc tgaccacccg cttccgcgtc accgccgaac 10440

gcgtcgaact gcgctcgggc ctgctcttcc gcagccgccg ctcggtcccc atcgaccggg 10500

tccgcagcgt cgacgtcgaa gccaagccgg tgcaccgcct cttcggcctc gcctcgctgc 10560

gcatcggcac cggtgaacag ggcgcgtcca gccgcaggct ctccctcgac ggcatcacca 10620

ggcgtcaggc gcggcgactg cgcaggctcc tcatcgaccg ccgtggcagc ggccatgcca 10680

ccggccagga ccaggacgtc accatcgccg agatggactg ggcctggctg cggtacgcgc 10740

cgctcaccat ctggggcgtc ggcagcgtct tcgccgccgt cggcaccgcc taccgcatcc 10800

tgcacgagat gaaggtcgac ccgctcgaac tgggcgtcgt caaggacatc gaggaccgct 10860

tcggttccgt acccctgtgg ttcggcatcc tcgtcgccgt cgtgatcacc gccgtcgtgg 10920

gcgccgcggt ctccaccgcc accttcgtgg acgcctggac caactaccgc ctggagcgtg 10980

aggggtcgg catcttccgg atccgccgcg gactgctcat ttcccgctcc gtcaccatcg 11040

aggagccgcg gctgcgcggc gtcgagctcg ccgagccgat gctgctgcgc tgggcgggcg 11100

gcgccaccct gagcgccatc gccagcggcc tcagcaacag ccaggagaac cgcagccgct 11160

gttccctcac cccgcccgtg ccccgggacg aggcgctgcg ggtcgccgcc gacgtcctcg 11220

ccgaggaagg gtccccgacg gagctgacca agctcgtccg gcactcccgt gccgccctgc 11280

gccgtcgcat caaccgcggc ctgctggtcc tcgcggccgt cgtcgcggtg ccgctgggcc 11340

tggggctgtg gctcaccccc gtcctggtgc acaccgcctg gatcacggcg ctcgtcggcc 11400

tgccggtcgt catcgtcctc gccaacgacg cctaccgctc cctcggccac ggaatccgcg 11460

accgctacct cgtcgtccgc gccggcacct tcgcccgccg tacggtcgcc gtccagcggg 11520

acggcgtcat cggctggaac atctcccgct cctacttcca gcggcgcagc ggactgctca 11580

ccatcggcgc caccaccgcg ggcgtcggct gccacaaggt gcgcgacgta tccgtcggcg 11640

ccggcctcgc cttcgccgaa gaggccgtac ccaggctgct cgccccgttc atcgaacgcg 11700

tcccgcgcgg ctgaaccccc tcagaccaac tggcgaaccc 11740

<210> 2
<211> 719
<212> PRT
<213> Streptomyces aizunensis

<400> 2

Met Asp Asn Leu Glu Leu Arg Arg Glu Ala Asp Ala Ile Leu Ala Glu
1 5 10 15

Leu Val Gly Ala Pro Gly Gly Ser Ala Arg Leu Arg Glu Asp Gln Trp
20 25 30

```
Gln Ala Val Ala Ala Leu Val Glu Glu Arg Arg Arg Ala Leu Val Val
        35                  40                  45

Gln Arg Thr Gly Trp Gly Lys Ser Ala Val Tyr Phe Val Ala Thr Ala
    50                  55                  60

Leu Leu Arg Arg Arg Gly Ser Gly Pro Thr Val Ile Ile Ser Pro Leu
65                  70                  75                  80

Leu Ala Leu Met Arg Asn Gln Val Glu Ala Ala Ala Arg Ala Gly Ile
                85                  90                  95

Gln Ala Arg Thr Ile Asn Ser Ala Asn Pro Glu Glu Trp Glu Thr Ile
            100                 105                 110

Tyr Gly Glu Val Glu Arg Gly Glu Thr Asp Val Leu Leu Val Ser Pro
        115                 120                 125

Glu Arg Leu Asn Ser Val Asp Phe Arg Asp Gln Val Leu Pro Lys Leu
    130                 135                 140

Ala Ala Thr Thr Gly Leu Leu Val Val Asp Glu Ala His Cys Ile Ser
145                 150                 155                 160

Asp Trp Gly His Asp Phe Arg Pro Asp Tyr Arg Arg Leu Arg Thr Met
                165                 170                 175

Leu Ala Glu Leu Pro Glu Gly Val Pro Val Leu Ala Thr Thr Ala Thr
            180                 185                 190

Ala Asn Ala Arg Val Thr Ala Asp Val Ala Glu Gln Leu Gly Thr His
        195                 200                 205

Gly Glu His Ala Leu Val Leu Arg Gly Pro Leu Asp Arg Glu Ser Leu
    210                 215                 220

Arg Leu Gly Val Leu Gln Leu Pro Asp Ala Ala His Arg Leu Ala Trp
225                 230                 235                 240

Leu Gly Asp Arg Leu Ala His Leu Pro Gly Ser Gly Ile Ile Tyr Thr
                245                 250                 255

Leu Thr Val Ala Ala Ala Glu Glu Val Ala Ala Phe Leu Arg Gln Arg
            260                 265                 270

Gly Tyr Pro Val Ala Ser Tyr Thr Gly Lys Thr Glu Asn Ala Asp Arg
        275                 280                 285

Leu Gln Ala Glu Glu Asp Leu Leu Ala Asn Arg Val Lys Ala Leu Val
    290                 295                 300

Ala Thr Ser Ala Leu Gly Met Gly Phe Asp Lys Pro Asp Leu Gly Phe
305                 310                 315                 320

Val Val His Met Gly Ser Pro Ser Ser Pro Ile Ala Tyr Tyr Gln Gln
                325                 330                 335

Val Gly Arg Ala Gly Arg Gly Val Asp His Ala Asp Val Leu Leu Leu
            340                 345                 350

Pro Gly Arg Glu Asp Glu Ala Ile Trp Ala Tyr Phe Ala Ser Val Gly
        355                 360                 365
```

```
Phe Pro Pro Glu Glu Gln Val Arg Arg Thr Leu Asp Val Leu Ala Gln
    370                 375                 380

Ala Gly Arg Pro Leu Ser Leu Pro Ala Leu Glu Pro Leu Val Asp Leu
385                 390                 395                 400

Arg Arg Ser Arg Leu Glu Thr Met Leu Lys Val Leu Asp Val Asp Gly
                405                 410                 415

Ala Val Lys Arg Val Lys Gly Gly Trp Thr Ala Thr Gly Gln Pro Trp
            420                 425                 430

Thr Tyr Asp Ala Glu Arg Tyr Ala Trp Val Ala Lys Gln Arg Ala Ala
        435                 440                 445

Glu Gln Gln Ala Met Arg Asp Tyr Val Ala Thr Thr Gly Cys Arg Met
    450                 455                 460

Glu Phe Leu Gln Arg Gln Leu Asp Asp Glu Lys Ala Val Pro Cys Gly
465                 470                 475                 480

Arg Cys Asp Asn Cys Ala Gly Ser Trp Leu Glu Ala Val Val Ser Pro
                485                 490                 495

Ala Ala Leu Ala Ala Ala Ala Gly Glu Leu Asp Arg Ala Gly Val Glu
            500                 505                 510

Val Glu Ser Arg Lys Met Trp Pro Thr Gly Leu Ala Ala Val Gly Met
        515                 520                 525

Asp Leu Lys Gly Arg Ile Pro Ala Gly Gln Gln Ala Val Thr Gly Arg
    530                 535                 540

Ala Leu Gly Arg Leu Ser Asp Ile Gly Trp Gly Asn Arg Leu Arg Pro
545                 550                 555                 560

Leu Leu Ser Ala Gln Ala Ala Asp Gly Pro Val Pro Asp Asp Val Leu
                565                 570                 575

Ala Ala Val Val Thr Val Leu Ala Asp Trp Ala Arg Ser Pro Gly Gly
            580                 585                 590

Trp Ala Ser Gly Gly Pro Asp Ala Met Ala Arg Pro Val Gly Ile Val
        595                 600                 605

Ala Met Pro Ser Arg Thr Arg Pro Arg Leu Val Ala Ser Leu Ala Glu
    610                 615                 620

Gly Val Ala Arg Val Gly Arg Leu Pro Leu Leu Gly Ser Leu Ala Tyr
625                 630                 635                 640

Thr Pro Gln Ala Asp Val Tyr Gly Ala His Arg Ser Asn Ser Ala Gln
                645                 650                 655

Arg Leu Arg Ala Leu Ala Asp Ser Phe Thr Val Pro Glu Glu Leu Ala
            660                 665                 670

Ala Ala Leu Ala Ala Ala Pro Gly Pro Val Leu Leu Val Asp Asp Tyr
        675                 680                 685

Thr Asp Ser Gly Trp Thr Leu Ala Val Gly Ala Arg Leu Leu Arg Gln
    690                 695                 700
```

Ser Gly Ala Gly Gly Val Leu Pro Leu Val Leu Ala Leu Ala Gly
705 710 715

<210> 3
<211> 2160
<212> DNA
<213> Streptomyces aizunensis

<400> 3
atggacaacc tggagctccg tcgtgaagcc gatgccatcc tcgctgagct ggtcggtgcc 60

cctgggggtt cggcgcggct gcgggaggac cagtggcagg cggtcgcggc cctggtggag 120

gagcgccggc gggccctggt ggtgcagcgc acgggctggg gcaagtccgc ggtctacttc 180

gtcgccaccg ctctgctgcg ccggcgcggc tccgggccga cggtgatcat ttctccgctg 240

ctggcgctga tgcgcaacca ggtcgaggcg gccgcgcggg ccgggatcca ggcgcgcacg 300

atcaactcgg ccaacccgga ggagtgggaa accatctacg gggaggtcga gcgcggcgag 360

accgatgtgc tcctcgtcag ccccgagcgc ctcaactccg tggatttccg cgaccaggta 420

ctgcccaagc tggcggccac gacgggtctg ctggtggtcg acgaggcgca ctgcatctcc 480

gactggggcc acgacttccg ccccgactac cgacggctgc gcacgatgct ggcggagctg 540

ccggagggcg tgccggtcct ggccacgacg gcgaccgcga acgcgcgggt gaccgcggac 600

gtggcggagc agctgggcac gcacggcgag cacgccctgg tcctgcgcgg accgctcgac 660

cgggagagcc tgcggctggg agtgctgcag ctgccggacg cggcgcaccg gctggcctgg 720

ctgggggacc ggctggcgca cctgccgggt tcggggatca tctacacgct gaccgtggcg 780

gcggcggagg aggtcgcggc gttcctgcgg caacgcgggt atccggtggc ttcctacacc 840

gggaagacgg agaacgccga ccggttgcag gcggaggagg atctgctggc gaaccgggtg 900

aaggcactgg tggcgacctc ggcgctgggc atggggttcg acaagccgga cctggggttc 960

gtggtgcaca tggggtcgcc ctcgtccccg atcgcctact accagcaggt ggggcgcgcg 1020

gggcgtgggg tggatcacgc ggacgtgctg ctgctgccgg gccgggagga cgaggcgatc 1080

tgggcgtact tcgcctcggt gggcttcccg cccgaggagc aggtccggcg caccctggac 1140

gtactggcgc aggcgggccg cccgctgtcg ctgcccgcgc tggagccgct ggtggacctc 1200

cggcgctcgc gcctggagac gatgctgaag gtcctggacg tggacggcgc ggtcaagcgc 1260

gtgaagggcg gctggaccgc caccgggcag ccgtggacgt acgacgcgga gcggtacgcc 1320

tgggtcgcga agcagcgggc ggcggagcag caggccatgc gggactacgt ggcgaccacg 1380

ggctgccgga tggagttcct gcagcggcag ctggacgacg agaaggcggt cccgtgcggc 1440

cgctgcgaca actgcgccgg atcctggctg gaggcggtcg tgtcgcccgc ggccctcgcg 1500

gccgcggcgg gcgagctgga ccgcgcgggg gtcgaggtcg agtcccgcaa gatgtggccg 1560

accgggctcg ccgcggtcgg catggacctg aagggccgga tccccgcggg ccagcaggcc 1620

gtcaccgggc gcgcgctcgg caggctgtcg gacatcggct ggggcaaccg gctgcgcccc 1680

ctgctgtcgg cgcaggccgc ggacgggccg gttccggacg atgtgctggc cgccgtcgtg 1740

acggtgctcg ccgactgggc ccgctcgccg ggcggctggg cgagcggcgg gccggacgcg 1800

atggcgcggc cggtggggat cgtcgccatg ccctcccgta cccgcccgcg gctggtcgcc 1860

tcgctggccg agggcgtggc ccgggtcggc aggctcccgc tgctgggcag cctcgcctac 1920

accccgcagg ccgacgtgta cggggcgcac cgcagcaact cagcccagcg gctgcgcgcc 1980

ctggccgact cgttcaccgt gcccgaggaa ctcgccgcgg ccctggccgc cgctcccggc 2040

ccggtcctgc tcgtcgacga ctacaccgac tccggctgga ccctggccgt gggcgcacgc 2100

ctgctgcgcc agtccggcgc gggcggcgtg ctcccgctcg tcctcgcgct ggccgggtag 2160

<210> 4
<211> 253
<212> PRT
<213> Streptomyces aizunensis

<400> 4

Leu Asn Lys Gln Glu Thr Ser Leu Trp Val Arg Arg Tyr His Ala Ser
1 5 10 15

Asp Glu Ser Arg Ile Gln Leu Val Cys Leu Pro His Ala Gly Gly Ser
20 25 30

Ala Ser Phe Tyr Phe Pro Met Ser Gln Ser Leu Ala Pro Ala Met Asp
35 40 45

Val Leu Ser Val Gln Tyr Pro Gly Arg Gln Asp Arg Arg Asp Glu Pro
50 55 60

Gly Ile Val Asp Ile Gly Ala Tyr Ala Asp Ala Leu Thr Glu Gln Leu
65 70 75 80

Val Pro Trp Leu Asp Arg Pro Leu Ala Phe Phe Gly His Ser Met Gly
85 90 95

Ala Ile Leu Ala Phe Glu Val Thr Arg Arg Leu Glu Arg Asp His Gly
100 105 110

Val Thr Pro Glu His Ile Phe Ala Ser Gly Arg Arg Ser Pro Ala Ser
115 120 125

Phe Arg His Glu Thr Val His Leu Arg Asp Asp Asp Gly Ile Val Ala
130 135 140

Glu Met Arg Glu Leu Ser Gly Thr Asp Ala Lys Ile Leu Gly Asn Glu
145 150 155 160

Glu Ile Leu Arg Met Val Leu Pro Ala Ile Arg Ser Asp Tyr Thr Ala
165 170 175

Ile Glu Asn Tyr Arg Ala Ala Pro Glu Asp Val Val Arg Thr Pro Ile
180 185 190

Thr Val Leu Thr Gly Asp Ala Asp Pro Arg Thr Ser Arg Glu Glu Ala

195 200 205

Asp Ala Trp Lys Ala His Thr Thr Gly Gly Phe Asp Leu His Ser Phe
210 215 220

Pro Gly Gly His Phe Phe Leu Ala Asn His Gln Glu Lys Ile Met Gly
225 230 235 240

Ile Ile Ser Glu Glu Leu Ser Ala Pro Ala Arg Met Ala
245 250

<210> 5
<211> 762
<212> DNA
<213> Streptomyces aizunensis

<400> 5
ttgaataagc aagaaactag cctctgggtt cgccgctacc acgcttcgga cgaaagccgg 60

atccaattgg tctgtctgcc gcacgccggt ggctcggcct ccttctactt ccccatgtcc 120

cagtcgctgg ctccggcgat ggacgtcctc tcggtccagt accccggcag gcaggaccgc 180

agggacgagc ccgggatcgt ggacatcggc gcctacgcgg acgccctgac cgagcaactc 240

gtaccgtggc tcgaccggcc cctggccttc ttcggccaca gcatgggtgc gatcctcgcc 300

ttcgaggtga cgcgcaggct ggagcgtgac cacggcgtca ctccggagca catcttcgct 360

tccggccggc gctcgcccgc cagtttccgg cacgagaccg tgcacctgcg ggacgacgac 420

ggaatcgtgg cggaaatgcg ggaactcagc ggaaccgacg cgaagatact cggcaacgag 480

gaaatcctcc gcatggtgct ccccgcgatt cgaagcgact acaccgccat cgagaactac 540

cgtgccgcgc cggaagacgt cgtgcgtact cccatcacgg tgctgaccgg tgacgcggac 600

ccgaggacca gccgggaaga ggcggacgcc tggaaggcgc acacgaccgg cggattcgat 660

ctgcattcct tccccggtgg acatttcttc ctggcgaatc accaggagaa gatcatggga 720

attatttcgg aggaactctc cgcgccggct cgcatggcgt ga 762

<210> 6
<211> 956
<212> PRT
<213> Streptomyces aizunensis

<400> 6

Val Ala Val Arg Leu Val Glu Arg Glu Lys Gln Leu Glu Thr Leu Lys
1 5 10 15

Glu Leu Leu Gly Ser Ala Val Arg Gly Arg Gly Arg Val Ala Val Ile
20 25 30

Ser Gly Ala Val Ala Gly Gly Lys Thr Ser Leu Leu Glu Ile Phe Thr
35 40 45

Glu Glu Ala Ile Ser Ala Gly Ala Leu Val Leu Glu Ala Thr Gly Ser
50 55 60

```
Arg Ala Glu Arg Tyr Leu Pro Phe Gly Ile Leu Arg Arg Ile Leu Asp
65                  70                  75                  80

Ser Ala Ala Pro Leu Ser Pro Glu Ile His Ala Tyr Ala Thr Glu Leu
                85                  90                  95

Leu Asp Arg Val Ser Ala Gly Thr Thr Asp Ala Glu Gly Ala Val Glu
            100                 105                 110

Ala Gly Met Arg Val Leu Pro His Val Ala Thr Ala Leu Leu Arg Ile
        115                 120                 125

Ala Arg Asn Arg Thr Val Val Ile Ala Ile Asp Asp Val His His Gly
    130                 135                 140

Asp Glu Leu Ser Leu Ala Phe Leu Leu Cys Leu Ala Arg Arg Val Arg
145                 150                 155                 160

Gln Ala Gly Val Leu Ile Val Leu Thr Glu Ala Val Arg Leu Arg Ser
                165                 170                 175

Ala Gln Leu Ala Phe His Ala Glu Leu Gln Arg Gln Pro Asn Cys Thr
            180                 185                 190

Ser Leu Arg Leu Pro Leu Leu Thr Thr Arg Gly Thr Thr Arg Val Leu
        195                 200                 205

Ala Glu His Phe Ser Pro Ser Thr Ala Gln Arg Leu Ser Ala Glu Cys
    210                 215                 220

Gln Glu Thr Thr Gly Gly Asn Pro Leu Leu Val Arg Ala Leu Ile Asp
225                 230                 235                 240

Asp Gly Leu Thr Ala Leu Gly Asp Ser Glu Pro Phe Gln Arg Leu Ala
                245                 250                 255

Pro Ala Glu Thr Phe Glu Arg Ala Val Leu Asp Cys Leu His Arg Gly
            260                 265                 270

Asp Pro Glu Leu Leu Thr Val Ala Arg Gly Val Ala Val Leu Gly Ser
        275                 280                 285

Ala Cys Ser Leu Ala Leu Leu Asn Gly Ile Val Asp Leu His Ala Lys
    290                 295                 300

Ala Thr Glu Gln Ala Leu Gln Asp Leu Ser Arg Cys Ala Val Leu His
305                 310                 315                 320

His Gly Ser Phe Arg Asp Pro Ala Ala Arg Thr Ala Val Leu Glu Ala
                325                 330                 335

Thr Pro Pro Ala Ala Leu Ser Ala Leu His Leu Arg Thr Ala Arg Leu
            340                 345                 350

Leu His Gln Glu Gly Ala Thr Ala Leu Asp Val Ala Arg His Leu Leu
        355                 360                 365

Ala Ala Arg Lys Asn Val Glu Asp Trp Ala Ile Pro Val Leu Gln Glu
    370                 375                 380

Ala Val Glu Tyr Ala Leu Val Glu Asp Glu His Glu Leu Ala Leu Arg
385                 390                 395                 400
```

```
Cys Gly Glu Leu Ala Val Ala Ser Cys Ala Glu Gly Pro Arg His Ala
                405                 410                 415

Ala Leu Lys Ser Arg Leu Ala Ser Ile Val Trp Arg Ser Ser Pro Ala
            420                 425                 430

Ala Ala Glu Gly His Leu Arg Gln Leu Ser Arg Glu Leu Ala Ala Gly
        435                 440                 445

Arg Leu Ala Asp Arg Asp Leu Val Gln Ala Val Ser Leu Leu Ala Trp
    450                 455                 460

Met Gly Glu Ser Arg Gly Ala Gly Glu Ala Val Leu Arg Leu Gln Arg
465                 470                 475                 480

Thr Asp Ser Glu Ala Glu Ala Ala Gly Arg Ala Pro Ala Tyr Asp Pro
                485                 490                 495

Gly Thr Leu Thr Ala Ala Gln Ser Trp Leu Ser Met Val Ser Pro Pro
            500                 505                 510

Ala Arg Asp Leu Phe Asp Ala Val Glu Pro Arg Arg Thr Thr Leu Ser
        515                 520                 525

Gly Ala Pro Gly Ala Leu Pro Gly Ala Gly Pro Asp Thr Val Pro Tyr
    530                 535                 540

Asp Met Pro Asp Asn Ala Tyr Val Gln Ala Ala Asp Ala Val Arg Thr
545                 550                 555                 560

Ala Leu Arg Gly Gly Thr Gln Ala Asp Ala Ala Val Ser Lys Ala Thr
                565                 570                 575

Arg Val Leu Gln Arg Tyr His Leu Ser Asp Arg Thr Leu Gln Pro Leu
            580                 585                 590

Val Phe Ala Leu Leu Ala Val Ile Tyr Ala Gly Arg Leu Asp Leu Ala
        595                 600                 605

Ser Ala Trp Cys Glu Arg Leu Leu Gly Glu Cys Ser Ala Arg Asn Ala
    610                 615                 620

Pro Thr Trp Gln Ala Ala Leu Gly Val Val Arg Ala Glu Ile Leu Leu
625                 630                 635                 640

Arg Gln Gly Asp Leu Pro Gly Ala Ala Ala Gln Ala Arg His Ala Met
                645                 650                 655

Ser Arg Ile Ser Leu Gln Ser Trp Gly Val Gly Ile Ala Leu Pro Leu
            660                 665                 670

Ala Val Leu Val Glu Ala Glu Val Gln Met Gly Asp His Glu Glu Ala
        675                 680                 685

Met Ser Leu Leu Glu Gln Pro Val Pro Gln Ala Met Phe Asp Thr Leu
    690                 695                 700

Ala Gly Leu His Tyr Leu Arg Ala Arg Gly Arg Cys His Leu Ala Thr
705                 710                 715                 720

Gly Arg Tyr His Ala Ala Val Arg Asp Phe Leu Asn Cys Gly Glu Leu
                725                 730                 735
```

```
Met Gln Ala Trp Gly Val Asp Gly Ala Glu Leu Val Pro Trp Arg Leu
            740                 745                 750

Asp Ala Ala Glu Ala Trp Leu Ala Leu Gly Asn Val Ala Arg Ala Lys
        755                 760                 765

Glu Tyr Thr Glu Gln Gln Lys Gln Arg Glu Thr Gly Pro Val Gly Ser
    770                 775                 780

Arg Thr Arg Gly Ser Leu Leu Leu Thr Leu Ala His Thr Gly Gly Asp
785                 790                 795                 800

Leu Thr Val Arg Leu Lys Arg Leu Val Glu Ala Val Glu Thr Leu Glu
                805                 810                 815

Glu Gly Gly Asp Arg Leu Gln Leu Ala Val Ala Leu Gly Glu Leu Gly
            820                 825                 830

Arg Gly Tyr Arg Ala Leu Gly Asp Phe Asn Arg Ala Arg Met Leu Val
        835                 840                 845

Arg Lys Ala Trp His Val Ala Lys Ser Cys Gly Ala Glu Pro Leu Cys
    850                 855                 860

Gln Gln Phe Met Pro Gly Gln Val Asp Gly Glu Ala Gly Ala Gln Ser
865                 870                 875                 880

Gly Arg Glu Ala Glu Leu Pro Ser Glu Val Glu Val Leu Ser Glu Ala
                885                 890                 895

Glu Ala Arg Val Ala Leu Leu Ala Ala Arg Gly His Thr Asn Arg Glu
            900                 905                 910

Ile Ala Thr Lys Leu Tyr Val Thr Val Ser Thr Val Glu Gln His Leu
        915                 920                 925

Thr Arg Ile Tyr Arg Lys Leu Lys Val Lys Arg Arg Arg Asp Leu Pro
    930                 935                 940

Ala Arg Leu Ser Asp Leu Ser Leu Pro Ser Ile Ala
945                 950                 955
```

<210> 7
<211> 2871
<212> DNA
<213> Streptomyces aizunensis

<400> 7

```
gtggcggtta ggctcgtcga gcgcgagaag cagctggaaa cgctgaagga actactcggc     60

agcgcagtcc gtggccgagg gcgggtcgcc gtcatcagcg gggcagtcgc cggcgggaaa    120

acgagtctgc tggaaatctt caccgaagag gcgatctccg cgggcgcgct ggtgctggaa    180

gccacgggct cccgggcgga gcgctatctg cccttcggaa ttctgcgcag aatcctcgac    240

agcgcggcgc ccctgtcgcc cgagatccac gcctacgcca ccgagctgct ggaccgcgtc    300

agcgccggga cgacggacgc cgaaggcgcc gtcgaggccg gtatgcgcgt cctgccccat    360

gtcgccaccg cactgttaag gatcgcccgg aaccggaccg tcgtcatagc catcgacgac    420

gtccaccacg gggacgaact ctccctcgcc ttcctgctgt gcctcgcccg ccgagtgcgc    480
```

caggcgggcg tcctgatcgt gctcaccgaa gccgtccggc tgcggtccgc gcaactcgcc 540

ttccacgccg aactgcagcg ccagcccaac tgcaccagcc tccggctgcc cctgctcacc 600

acgcgcggca ccacccgcgt cctcgccgag cacttctccc cctcgacggc gcaacggctg 660

tccgccgagt gccaggagac caccggcggc aatccactgc tggtcagggc gctgatcgac 720

gacggcctca cggcgctcgg agacagcgag cccttccagc ggctcgcccc cgccgaaacc 780

ttcgaacgcg ccgtgctcga ctgcctgcac cgcggcgacc ccgagctgct gaccgtcgcc 840

cggggcgtcg ccgtactcgg tagcgcctgc tccttggccc tgctcaacgg gatcgtcgac 900

ctgcacgcca aggccaccga acaggccctt caggacctca gccggtgcgc cgtcctgcac 960

cacggctcct tccgcgaccc ggcggcccgt accgccgtcc tggaagccac tccgcccgcg 1020

gcgctgtccg ccctgcacct gcgcaccgcg cgactcctgc accaggaagg cgcgacggcg 1080

ctcgatgtcg cccgccacct cctcgccgcc cgcaagaacg tcgaggactg ggcgatcccc 1140

gtcctccagg aggcggtcga gtacgccctc gtcgaggacg agcacgaact cgccctgcgg 1200

tgcggggaac tggcggtcgc ctcctgcgcg gagggccccc gacacgccgc cctgaagtcc 1260

cgcctggcga gcatcgtctg gcgcagcagc ccggccgccg ctgaagggca tctgcggcag 1320

ctgtcccgcg aactcgccgc cggccggctc gccgaccgcg atctcgtcca ggccgtgtcg 1380

ctcctggcgt ggatggggga gtcccggggg gccggcgagg cggtactgcg actgcagcgg 1440

accgacagcg aggccgaggc ggccggacgg gcgcccgcct acgacccggg cacgctcacc 1500

gccgcacaga gctggctctc gatggtcagc ccgccggccc gcgacctctt cgacgccgtg 1560

gaaccgcgcc ggacaacgct gtcaggcgcg ccgggggcgc tgcccggcgc ggggcccgac 1620

accgtcccct acgacatgcc cgacaacgcc tacgtccagg ccgccgacgc cgtccgcacc 1680

gccctgcgcg gcggaaccca ggccgacgcc gccgtcagca aggccacccg ggtgctccag 1740

cgctaccacc tgagcgaccg cacctccag ccgctcgtct tcgccctcct cgccgtcatc 1800

tacgcgggtc gcctcgacct cgcgtccgcc tggtgcgaac gactgctcgg cgagtgctcc 1860

gcccgcaacg ccccgacctg gcaggccgcc ctcggtgtgg tccgggccga gatcctgctg 1920

cgccagggcg atctgcccgg tgcggccgcc caggcccgcc acgccatgtc ccggatctcc 1980

ctgcagagct ggggcgtggg catcgcgctg ccgctggccg tcctcgtcga ggccgaggtc 2040

cagatgggcg accacgagga ggcgatgagc ctgctcgaac agccggtgcc ccaggccatg 2100

ttcgacaccc tggccggcct gcactacctc agggcccgcg gccgctgcca cctggccacc 2160

ggccgctacc acgccgccgt gcgggacttc ctgaactgcg gcgagctgat gcaggcctgg 2220

ggcgtggacg gggcggagct ggtgccgtgg cggctggacg ccgccgaggc gtggctggcc 2280

ctcggcaacg tcgcgcgcgc caaggagtac accgagcagc agaagcagcg cgagacgggg 2340

cccgtgggca gccggacgcg tggctccctg ctgctcacgc tcgcccacac cggcggtgac 2400

ctcacggtcc ggctcaagcg gctcgtcgag gccgtcgaga ccctggagga gggcggggac 2460

cggctccagc tggcggtggc gctgggggag ctgggccgcg gctaccgtgc gctgggcgac 2520

ttcaaccggg cccggatgct ggtgcgcaag gcctggcacg tcgccaagtc ctgcggcgcc 2580

gaaccgctgt gccagcagtt catgccgggg caggtcgacg gcgaggccgg tgcgcagagc 2640

ggccgggagg cggagcttcc cagcgaggtc gaggtcctgt ccgaggccga ggcgcgggtc 2700

gcgctgctgg cggcgcgcgg ccacaccaac cgtgagatag cgaccaagct ctacgtcacg 2760

gtgtccacgg tcgagcagca tctgacgcgc atctaccgca agctgaaggt gaagcggcgc 2820

cgcgatctgc ccgcccggct gtcggacctg agcctgccga gcatcgcctg a 2871

<210> 8
<211> 201
<212> PRT
<213> Streptomyces aizunensis

<400> 8

Met Leu Val Asp Asp His Pro Val Val Arg Glu Gly Leu Ser Ser Met
1 5 10 15

Leu Gln Ser Ala Asp Gly Val Ser Val Val Gly Gln Ala Asp Ser Gly
20 25 30

Glu Glu Ala Ile Ala Met Val Thr Arg Leu Ala Pro Asp Ile Val Leu
35 40 45

Leu Asp Leu Gln Met Gly Gly Met Asp Gly Val Glu Thr Thr Gly His
50 55 60

Leu Leu Arg Val Ala Pro Ala Thr Lys Val Val Ile Val Thr Thr Tyr
65 70 75 80

Glu Ser Asp Thr Asp Ile Leu Arg Ala Val Glu Ala Gly Ala Ala Gly
85 90 95

Tyr Leu Leu Lys Gly Ser Ser Arg Asp Glu Leu Val Gln Ala Val Lys
100 105 110

Ala Ala Ala Arg Gly Glu Thr Val Leu Thr Pro Ser Leu Ala Pro Lys
115 120 125

Leu Phe Arg Ala Arg Val Val Glu Pro Pro Ala Leu Ser Asp Arg Glu
130 135 140

Arg Glu Val Leu Gln Leu Val Ser Leu Gly Leu Thr Asn Ala Asp Ile
145 150 155 160

Gly Arg Gln Leu Phe Ile Ser Glu Ala Thr Val Lys Thr His Leu Leu
165 170 175

Arg Ser Phe Lys Lys Leu Ser Val Ser Asp Arg Thr Ala Ala Val Ile
180 185 190

Thr Ala Leu Lys Leu Gly Leu Leu Ser

195 200

<210> 9
<211> 606
<212> DNA
<213> Streptomyces aizunensis

<400> 9
atgctggtcg acgaccaccc ggtcgtacgg gagggtctca gctcgatgct gcaatccgcc 60

gacggcgtga gcgtcgtcgg gcaggccgac tcgggcgagg aggcgatcgc gatggtcacc 120

agacttgctc ccgacatcgt cctgctcgac cttcagatgg gcgggatgga cggggtggag 180

accaccggcc acctgctgcg cgtcgcgccg gccaccaagg tggtgatcgt gacgacgtac 240

gagagcgaca cggacatcct gcgggccgtg gaggcgggcg cggcgggcta cctgctcaag 300

ggcagctcgc gcgacgaact ggtgcaggcg gtaaaggcgg cggcccgcgg tgagacggtc 360

ctgacgccgt cgctcgcccc caagctgttc cgggcccggg tggtggagcc gcccgcgctg 420

tcggaccgtg agcgcgaggt gctccagttg gtcagcctcg ggctgaccaa cgcggacatc 480

ggccgccagc tgttcatcag cgaggcgacg gtgaagacgc atctgctgcg ctcgttcaag 540

aagctgtcgg tctcggaccg gacggccgcg gtgatcacgg cactgaagct cggcctgctg 600

tcctga 606

<210> 10
<211> 416
<212> PRT
<213> Streptomyces aizunensis

<400> 10

Val Ser Thr Glu Ser Pro Ala Ala Gln Ala Thr Asp Gly Gln Asp Asp
1 5 10 15

Ala Pro Glu Ala Gly Ala Arg Trp Phe Gly Leu Trp Asp Ala Leu Phe
20 25 30

Ala Val Ser Tyr Ala Val Thr Thr Met Leu Leu Phe Thr Ser Asp Gly
35 40 45

Glu Gln Val His Arg Ala Val Ala Met Ala Ala Leu Thr Ala Ala Val
50 55 60

Pro Trp Tyr Ala Ala Leu Gly Arg Ser Thr Met Val His Glu Gly Gln
65 70 75 80

Gly Pro Val Arg Arg Ser Val Val Phe Ser Val Gly Leu Phe Val Leu
85 90 95

Phe Ala Val Ala Val Val Phe Asp Leu Ala Ala Ser Phe Ala Leu Phe
100 105 110

Ala Val Val Pro Met Leu Met Met Ser Leu Ala Thr Ser Pro Ala Ile
115 120 125

Ala Val Val Thr Leu Ala Asn Leu Val Pro Val Ile Val Val Trp Leu

130 135 140

Arg Gly Gly Thr Leu Ser Arg Asp Ile Leu Ala Val Leu Pro Thr Ser
145 150 155 160

Leu Leu Gly Ile Ala Leu Ser Val Met Leu Gly Leu Trp Ile Thr Arg
165 170 175

Val Thr Arg Gln Ser Arg Asp Arg Ala Glu Leu Ile Glu Glu Leu His
180 185 190

Arg Asn Arg Ala Gln Val Ala Arg Leu Ser Arg Lys Ala Gly Val Ser
195 200 205

Ala Glu Arg Glu Arg Leu Ala Arg Glu Ile His Asp Thr Leu Ala Gln
210 215 220

Gly Leu Thr Ser Ile Ile Ser Leu Val Gln Ala Ala Glu Thr Asp Phe
225 230 235 240

Thr Ala Asp Pro Asp Arg Ala Arg Ala His Leu Ala Leu Ala Gly Arg
245 250 255

Val Ala Arg Glu Ser Leu Gly Glu Ala Arg Glu Phe Val Thr Glu Leu
260 265 270

Thr Pro Pro Ala Leu Gln Glu Ser Ser Leu Val Gln Ala Thr Arg Arg
275 280 285

Gln Ala Glu Gly Leu Thr Ala Gln Thr Gly Met Arg Ala His Val Thr
290 295 300

Val Glu Gly Asp Glu Arg Glu Leu Pro Met Ala Ile Ser Val Val Leu
305 310 315 320

Leu Arg Ser Leu Gln Glu Ala Ile Ala Asn Ile Arg Lys His Ala Gly
325 330 335

Lys Ala Arg Ala Ala Glu Ile Arg Leu Val Tyr Glu Gln Asp Thr Val
340 345 350

Arg Leu Leu Val Arg Asp Asp Gly Pro Gly Phe Thr Val Thr Gly Asp
355 360 365

Gln Arg Gly Ser Gly Leu Arg Gly Met Gln Thr Arg Ala His Glu Ile
370 375 380

Ser Gly Ala Ala Thr Val Val Ser Ser Pro Gly Gln Gly Thr Thr Ile
385 390 395 400

Glu Val Thr Val Pro Val Pro Ala Lys Gly Glu Glu Ala Asp Glu Arg
405 410 415

<210> 11
<211> 1251
<212> DNA
<213> Streptomyces aizunensis

<400> 11
gtgagcactg aatcccccgc ggcgcaggcg acagacggcc aggacgacgc gcccgaggcg 60

ggagcccggt ggttcggcct gtgggacgcg ctcttcgcgg tctcgtacgc cgtcaccacc 120

atgctgctgt tcacctccga cggtgaacag gtccaccggg ccgtggccat ggccgcgctg 180

accgcggccg tgccctggta cgcggccctg gggcgcagca ccatggtcca cgagggccag 240

gggcccgtcc ggcgcagcgt cgtcttctcc gtcggcctgt tcgtgctgtt cgcggtggcc 300

gtggtcttcg acctcgcggc ctcgttcgcg ctgttcgccg tggtcccgat gctgatgatg 360

agcctggcga cctcgccggc catcgccgtg gtcacgctgg ccaatctggt tccggtcatc 420

gtggtgtggc tgcgcggcgg caccctgagc cgcgacatcc tggccgtcct gccgacctcg 480

ctcctcggca tcgccctgtc ggtcatgctc gggctgtgga tcacccgggt gacccggcag 540

agccgtgacc gggccgagct catcgaggag ttgcaccgca accgtgcgca agtcgcccgg 600

ctgtcgcgca aggcgggggt ctccgccgag cgcgagcggc tcgcccggga gatccacgac 660

accctcgccc agggcctcac cagcatcatc agcctcgtac aggccgccga gaccgacttc 720

acggccgacc cggaccgggc gagggcgcat ctggcactgg cgggccgcgt ggcccgcgaa 780

agcctgggag aagcccgcga gttcgtcacc gagctgaccc cgcccgcgct gcaggagtcc 840

tcgctcgtgc aggcgacgcg gcggcaggcc gagggcctga cggcgcagac cggcatgcgg 900

gcgcacgtca ccgtcgaagg agacgagcgg gaactgccga tggcgatcag cgtggtcctg 960

ctgcgttccc tccaggaggc catcgcgaac atccgcaagc acgcggggaa ggcacgcgcg 1020

gccgagatcc ggctcgtgta cgaacaggac accgtacggc tgctcgtacg ggacgacgga 1080

cccgggttca ccgtcaccgg ggaccagcgg ggaagcgggc tgcgcggcat gcagactcgc 1140

gcacacgaga tcagcggggc ggcgaccgtg gtcagcagcc ccggacaggg caccaccatc 1200

gaagtgacgg tgcccgtgcc cgcgaaggga gaggaagccg atgagcgctg a 1251

<210> 12
<211> 186
<212> PRT
<213> Streptomyces aizunensis

<400> 12

Leu Ser Pro Phe Leu Asn Thr Leu Ile Ala Ser Gly Thr Ile Leu Ala
1 5 10 15

Val Ile Leu Ser Thr Asp Leu Gly Thr Arg Lys Val Thr Thr Thr Arg
20 25 30

Met Leu Pro Ser Leu Leu Ala Val Val Val Ile Leu Ala Leu Leu Val
35 40 45

His Thr Leu Pro Leu Asp Gly Asn Asp Pro Ser Leu Gln Leu Ala Gly
50 55 60

Ile Gly Ala Gly Ile Ile Cys Gly Leu Ala Ala Thr Ala Leu Leu Pro
65 70 75 80

Ala His Arg Asn Ala Ser Gly Glu Val Ser Thr Lys Gly Gly Ile Gly
85 90 95

```
Tyr Ala Leu Val Trp Thr Ala Leu Ser Ala Ser Arg Val Leu Phe Ala
            100                 105                 110

Tyr Gly Ser Gln His Trp Phe Ser Glu Gly Ile Val Arg Phe Ser Thr
        115                 120                 125

Asp Tyr Lys Leu Ser Gly Gln Ala Val Tyr Ser Asn Ala Phe Ala Phe
    130                 135                 140

Met Ala Leu Ala Met Val Leu Thr Arg Thr Ala Val Leu Leu Asn Thr
145                 150                 155                 160

Arg Arg Arg Leu Arg Gly Gly Gln Leu Pro Ala Ala Asp Asn Thr Ala
                165                 170                 175

Pro His Gln Ala Ser Ser Ala Asn Thr His
            180                 185
```

<210> 13
<211> 561
<212> DNA
<213> Streptomyces aizunensis

<400> 13

```
ctgagcccgt tcttgaacac actgatcgcc agcgggacga tcttggccgt cattctgtcg     60

accgacctcg gcacccgcaa agtcaccacg acgcggatgc ttccttcgct cctcgcggtc    120

gtcgtgatcc tcgcgctcct cgtgcacaca ctgccgctcg acggcaacga cccctcgctc    180

caactggcgg gcatcggcgc cggtatcatc tgcggactgg ccgccacggc gctcctcccc    240

gcccaccgga acgcttccgg tgaggtctcc accaagggcg gtatcggtta cgcgctggtg    300

tggaccgcgc tgtccgcctc gcgtgtgctc ttcgcctacg gttcacagca ctggttcagc    360

gagggcatcg tccggttcag caccgactac aagctcagcg gacaggccgt ctactccaac    420

gctttcgcct tcatggccct ggccatggtg ctgacgcgga ccgccgtcct gttgaacacg    480

cgccgccggc tgcgcggcgg gcagcttccc gcggccgaca acacggcccc acatcaggcg    540

agttccgcca atacgcactg a                                              561
```

<210> 14
<211> 163
<212> PRT
<213> Streptomyces aizunensis

<400> 14

```
Met His Asp Val Arg Leu Arg Pro Pro Arg Asn Arg Val Asp Ser Arg
1               5                   10                  15

Ala Val Gly Trp Trp Thr Val Gln Ser Ala Met Tyr Ala Leu Pro Leu
            20                  25                  30

Pro Ile Thr Phe Gly Val Leu Tyr Leu Cys Ile Pro Pro Ala Arg Pro
        35                  40                  45

Phe Phe Gly Trp Ala Phe Leu Ile Ser Leu Val Pro Gly Leu Ala Tyr
```

```
    50                  55                  60

Met Ala Val Met Pro Ala Trp Arg Tyr Arg Val His Arg Trp Glu Thr
65                  70                  75                  80

Thr Asp Glu Ala Val Tyr Ala Ala Ser Gly Trp Leu Trp Gln Gln Trp
                85                  90                  95

Arg Val Val Pro Met Ser Arg Ile Gln Thr Val Asp Thr Leu Arg Gly
            100                 105                 110

Pro Leu Gln Gln Leu Phe Gly Leu Ser Gly Ile Thr Val Thr Thr Ala
        115                 120                 125

Ser Tyr Ser Gly Ala Val Lys Ile Lys Gly Ile Asp His Arg Thr Ala
    130                 135                 140

Arg Asp Val Val Glu His Leu Thr Arg Val Thr Gln Ala Thr Pro Gly
145                 150                 155                 160

Asp Ala Thr
```

<210> 15
<211> 492
<212> DNA
<213> Streptomyces aizunensis

<400> 15
atgcatgatg tcaggctccg gcccccgcgc aatcgtgtcg actcccgggc agtgggctgg 60

tggacggtcc agtccgcgat gtacgccctg cccctgccga tcaccttcgg cgtgctgtac 120

ctgtgcatcc cgcccgccag gccgttcttc ggctgggcct tcctgatctc gctcgtaccg 180

ggcctcgcct acatggccgt catgcccgcc tggcgctacc gggtgcaccg ttgggagacc 240

accgacgaag ccgtctacgc ggcgtccggc tggctctggc agcagtggcg ggtcgtgccg 300

atgtcccgca tccagacggt ggacaccctg cgcggacccc tccagcagct cttcggcctc 360

tccggcatca ccgtcaccac cgcctcctac tccggcgccg tgaagatcaa gggaatcgac 420

caccggaccg cgcgggacgt ggtcgagcac ctcaccaggg tgacccaggc cacccccgga 480

gacgcgacat ga 492

<210> 16
<211> 514
<212> PRT
<213> Streptomyces aizunensis

<400> 16

```
Met Ser His Asp Thr Gly Gln Trp Glu Ala Thr Ala Thr Ser His Gly
1               5                   10                  15

Ala Ala Glu Asp Pro Glu Trp Ser Arg Leu Ser Pro Arg Leu Leu Leu
            20                  25                  30

Val Asn Leu Ser Met Leu Ala Gly Pro Leu Ala Leu Phe Ala Val Thr
        35                  40                  45
```

Val Ala Leu Thr Gly Ala Asn Leu Gln Ala Leu Ile Ser Leu Gly Ser
50 55 60

Leu Leu Ile Val Phe Leu Val Ile Thr Gly Ile Ser Thr Met Arg Leu
65 70 75 80

Leu Thr Thr Arg Phe Arg Val Thr Ala Glu Arg Val Glu Leu Arg Ser
85 90 95

Gly Leu Leu Phe Arg Ser Arg Arg Ser Val Pro Ile Asp Arg Val Arg
100 105 110

Ser Val Asp Val Glu Ala Lys Pro Val His Arg Leu Phe Gly Leu Ala
115 120 125

Ser Leu Arg Ile Gly Thr Gly Glu Gln Gly Ala Ser Ser Arg Arg Leu
130 135 140

Ser Leu Asp Gly Ile Thr Arg Arg Gln Ala Arg Arg Leu Arg Arg Leu
145 150 155 160

Leu Ile Asp Arg Arg Gly Ser Gly His Ala Thr Gly Gln Asp Gln Asp
165 170 175

Val Thr Ile Ala Glu Met Asp Trp Ala Trp Leu Arg Tyr Ala Pro Leu
180 185 190

Thr Ile Trp Gly Val Gly Ser Val Phe Ala Ala Val Gly Thr Ala Tyr
195 200 205

Arg Ile Leu His Glu Met Lys Val Asp Pro Leu Glu Leu Gly Val Val
210 215 220

Lys Asp Ile Glu Asp Arg Phe Gly Ser Val Pro Leu Trp Phe Gly Ile
225 230 235 240

Leu Val Ala Val Val Ile Thr Ala Val Val Gly Ala Ala Val Ser Thr
245 250 255

Ala Thr Phe Val Asp Ala Trp Thr Asn Tyr Arg Leu Glu Arg Glu Gly
260 265 270

Val Gly Ile Phe Arg Ile Arg Arg Gly Leu Leu Ile Ser Arg Ser Val
275 280 285

Thr Ile Glu Glu Arg Arg Leu Arg Gly Val Glu Leu Ala Glu Pro Met
290 295 300

Leu Leu Arg Trp Ala Gly Gly Ala Thr Leu Ser Ala Ile Ala Ser Gly
305 310 315 320

Leu Ser Asn Ser Gln Glu Asn Arg Ser Arg Cys Ser Leu Thr Pro Pro
325 330 335

Val Pro Arg Asp Glu Ala Leu Arg Val Ala Ala Asp Val Leu Ala Glu
340 345 350

Glu Gly Ser Pro Thr Glu Leu Thr Lys Leu Val Arg His Ser Arg Ala
355 360 365

Ala Leu Arg Arg Arg Ile Asn Arg Gly Leu Leu Val Leu Ala Ala Val
370 375 380

```
Val Ala Val Pro Leu Gly Leu Gly Leu Trp Leu Thr Pro Val Leu Val
385                 390                 395                 400

His Thr Ala Trp Ile Thr Ala Leu Val Gly Leu Pro Val Val Ile Val
                405                 410                 415

Leu Ala Asn Asp Ala Tyr Arg Ser Leu Gly His Gly Ile Arg Asp Arg
            420                 425                 430

Tyr Leu Val Val Arg Ala Gly Thr Phe Ala Arg Arg Thr Val Ala Val
        435                 440                 445

Gln Arg Asp Gly Val Ile Gly Trp Asn Ile Ser Arg Ser Tyr Phe Gln
    450                 455                 460

Arg Arg Ser Gly Leu Leu Thr Ile Gly Ala Thr Thr Ala Gly Val Gly
465                 470                 475                 480

Cys His Lys Val Arg Asp Val Ser Val Gly Ala Gly Leu Ala Phe Ala
                485                 490                 495

Glu Glu Ala Val Pro Arg Leu Leu Ala Pro Phe Ile Glu Arg Val Pro
            500                 505                 510

Arg Gly
```

<210> 17
<211> 1545
<212> DNA
<213> Streptomyces aizunensis

<400> 17

```
atgagccacg acaccggaca gtgggaggcc accgcgacct cccacggcgc cgccgaagac     60

cccgagtgga gcaggctcag cccccgactg ctgctggtca acctgagcat gctcgccggc    120

ccgctcgccc tgttcgccgt cacggtcgcc ctgaccggcg ccaacctcca ggccctcatc    180

tccctcggct ccctgctgat cgtcttcctg gtcatcaccg ggatcagcac gatgcggctg    240

ctgaccaccc gcttccgcgt caccgccgaa cgcgtcgaac tgcgctcggg cctgctcttc    300

cgcagccgcc gctcggtccc catcgaccgg gtccgcagcg tcgacgtcga agccaagccg    360

gtgcaccgcc tcttcggcct cgcctcgctg cgcatcggca ccggtgaaca gggcgcgtcc    420

agccgcaggc tctccctcga cggcatcacc aggcgtcagg cgcggcgact gcgcaggctc    480

ctcatcgacc gccgtggcag cggccatgcc accggccagg accaggacgt caccatcgcc    540

gagatggact gggcctggct gcggtacgcg ccgctcacca tctggggcgt cggcagcgtc    600

ttcgccgccg tcggcaccgc ctaccgcatc ctgcacgaga tgaaggtcga cccgctcgaa    660

ctgggcgtcg tcaaggacat cgaggaccgc ttcggttccg tacccctgtg gttcggcatc    720

ctcgtcgccg tcgtgatcac cgccgtcgtg ggcgccgcgg tctccaccgc caccttcgtg    780

gacgcctgga ccaactaccg cctggagcgt gagggggtcg gcatcttccg gatccgccgc    840

ggactgctca tttcccgctc cgtcaccatc gaggagcgcc ggctgcgcgg cgtcgagctc    900
```

gccgagccga tgctgctgcg ctgggcgggc ggcgccaccc tgagcgccat cgccagcggc 960

ctcagcaaca gccaggagaa ccgcagccgc tgttccctca ccccgcccgt gccccgggac 1020

gaggcgctgc gggtcgccgc cgacgtcctc gccgaggaag ggtccccgac ggagctgacc 1080

aagctcgtcc ggcactcccg tgccgccctg cgccgtcgca tcaaccgcgg cctgctggtc 1140

ctcgcggccg tcgtcgcggt gccgctgggc ctggggctgt ggctcacccc cgtcctggtg 1200

cacaccgcct ggatcacggc gctcgtcggc ctgccggtcg tcatcgtcct cgccaacgac 1260

gcctaccgct ccctcggcca cggaatccgc gaccgctacc tcgtcgtccg cgccggcacc 1320

ttcgcccgcc gtacggtcgc cgtccagcgg gacggcgtca tcggctggaa catctcccgc 1380

tcctacttcc agcggcgcag cggactgctc accatcggcg ccaccaccgc gggcgtcggc 1440

tgccacaagg tgcgcgacgt atccgtcggc gccggcctcg ccttcgccga agaggccgta 1500

cccaggctgc tcgccccgtt catcgaacgc gtcccgcgcg gctga 1545

<210> 18
<211> 164051
<212> DNA
<213> Streptomyces aizunensis

<400> 18
ctggctcagc ccgccagctc ctccagcctc ggcaccagcg acaccggaga gggcatcgtc 60

cggatctccg cgcgcacctc gcgcgcggcc gccgtcatct tctcgtccga aagcagctgt 120

acgaggacct ccgcggagag gtcgtcggcg gtgccgagca gaccggcacc ccggtcccgt 180

acggcctccg cattgatgtg acggtccgct ccgtccggca ggacgagctg cggcacaccg 240

gcgttcagcg ccgccagcgt cgtccccgca ccaccgtggt gcacggccgc gtcgcaggtc 300

tgcagcagcg ccgtcagcgg cacccacccc acggcccgga cgttgggagg cagttcaccg 360

agcgccgtgg tgtccacatc gcccagcgcc agcacgaact cggcgtccac cccggcagcc 420

gccgccgcga gccgctgcac cgggcccagg ccgttgatgt ggaccgaggc cgtgccgagc 480

gtcaccccga cccggcggcg ccccggcttc tccagcagcc agtccggcag caccgcaccg 540

ctgttgtacg ggaccggccg catcgaccag ccgtcccgct cgggctccgc catgctcggc 600

ggcgcgatgt cgatcaccgg gacccgttcg gacacccggt ccacgccgtg ccgcgccatc 660

gtctcggtga gcatcgacac cgtcagctcg cgcagctgcg taccccgcgc gaaaccgaag 720

ttgtgctgca cggccggcac acccagccgc gccgccgcga tcagaccgga cacgaagatc 780

tgctcgaaga cgatcagatc gggccggaaa tcgtcggcgg tccgcacgat gccgtccgcc 840

aggtggttgt tgaggtgggc gaagagggtc agcccgtcca tcgggtcgac gccgcccgga 900

ccgcgcaggc gggccatcag ctcaccggcc gtcgactgga ggaagtcctc caggtggaag 960

ccgggggcga catccgccac gtgcagaccg gcgttggcgg cctccagcgc gtcacccgcg 1020

ctggcgacca gcacctcgtg gccggccgag cgcaacgccc aggccagcgg aacaatggga 1080

aaaacgtggc cgatggccgg atacgtcacg aacagtatgc gcaaggaaac gcgccccctt 1140

gggtagcttt gtattctccg gaccggtatg gtccagatgg aatacggtgg atattcttta 1200

aatccccgac ggtgcctggg catcctgatg cagtcgcaca tgccgagtca aggcggcgtc 1260

cgaaggcccg tgttaggggt ccgtaggggc ctgttagggg tttctcccac ttccctcgca 1320

tgcaagagtg tcccctggtc ttggattctt tattcggggg taatggagcg cgcgatgttg 1380

aatgagtccg aggaattcac gcccgaaatc aatgtcgcct ccgaagtcgg tggaacgcag 1440

ggcgaaagtc ctgaaagcac gccgtcgtgg cagcagcgcc tgaccggcct caccgaggcc 1500

gagcagcaca ccgcactgct ggagtgggtg tcctcgctgg catccgccgc actgcgcgac 1560

gcggcccccg acacgctcga cccccaccgc cccttcctgg atctgggctt cgactcgctc 1620

gccgccgtcg acctgcacgc caggctcgtc gcgggaaccg ggctgcggct gccggtcacc 1680

ctggccttcg accaccccac ccccgcgcac ctcgcccgtc atctgcacgc ggcgatcctc 1740

ggactgaccg gccccgccga gacgcccgtc accgcggcgg tcggcagcga cgaacccatc 1800

gccatcgtcg gcatcggctg ccatttcccg ggcggcgtac agtcccccga ggcgctgtgg 1860

aacctcgtcg agaccggcac cgacgccatt tccgcattcc ccaccgggcg cggctgggat 1920

ctcgacgcgc tgtatgaccc ggatcccgac cgggcgggca ccagttatgc ccgcgagggc 1980

ggattcctgc acgacgccga cgcattcgac gcggcattct tcgggatatc cccgcgcgaa 2040

gccctcgcca tggatccgca gcagcgactc cttctcgaag cgtcctggga ggcattcgac 2100

cgcgccgggg tagaccccgc cgcattgcgc ggcggtcagg tcggcgtatt cgtcggcgcc 2160

gagacccagg aatacggccc ccggctccag gacgccaccg acggattcga gggctacctc 2220

gtcaccggaa acgcggccag cgtcgcctcc ggccgtatcg cctacacctt cggcttcgag 2280

ggcccgacgg tcaccgtcga cacggcctgc tcctcctcac tcgccgccct ccacctcgcc 2340

gtccaggcgc tgcgcaccgg cgaatgctcc ctcgcgctcg ccggtggcgt cgcggtcatg 2400

gcgagccccg gctcgttcgt ctcgttcagc cgccagcgcg gcctggcccc cgacggccgc 2460

tgcaagccgt tcgcggccgc cgccgacggc acggcgtggg gcgagggcgt cggcatgctg 2520

ctggtcgaac ggctctccga cgcgcgcgcc aagggccacc ggatcctcgc ggtcgtccgc 2580

ggctccgcca tcaaccagga cggcgccagc aacggcctca ccgcccccag cggtccgtcc 2640

cagcagcgcg tcatccgcca ggccctcgcc aacgccggcc tgtccgccgc cgaggtcgac 2700

gtcgtcgagg cgcacggcac cggcacccgg ctcggcgacc cgatcgaggc ccaggcgctc 2760

ctcgccacgt acggccagga gcacaccgat gaccggccgc tgtggctcgg ctccctgaag 2820

tcgaacatcg gccacacgca ggccgccgcc ggagtcgccg gcatcatcaa gatgatcatg 2880

gcgatgcggc acggggtact gccccggacc ctgcacgtcg acgcgccgac cccgcacgtc 2940

gactgggagg ccggagcggt caccttgctg accgaagccg tggagtggcc ggagtcggac 3000

cgcccgcgcc gtgcgggcgt gtcctccttc ggcatgagcg gcaccaacgc ccacgtcatc 3060

gtcgaagagc cggccgccca ggaccgcgag ggcgccccca cctccggcgc ccaagccccc 3120

gactccagcc agggccaggc acagggcacc tccaccgcgc cggttctcct cccgtgggcg 3180

ctctccgcca agacccccga ggccctccgc gcccaggcac gccgactcgg caccctgatc 3240

gcggcgcagc cgcacgtcac ccccctcgac atcggccact ccctcgcgac cacccggggc 3300

cgcttcgagc agcgcgccat cgtgctcggc gacgaccgcg aggcgttcct cgacgccctg 3360

cacgccctcg ccgagggcaa cgacacgccc tccgtggtcc agggcgccgc cgcaccgggc 3420

aagctcgcct tcctcttcac cggccagggc agccagcgcc tcggcatggg ccgcgaactg 3480

tacgagaccc acccggtgtt cgccgacgcc ctcgacgacg cctgctggta cctggacgac 3540

caactcgaac tcccgctcct cgacgtgctg ttcgccgacg agggcagccc cgaggccgca 3600

cttctgcacc agaccgccta cacgcagccc gcgctgttcg cggtcgaggt ggcgctgttc 3660

cgcctggtcg acagctgggg cctgaagccc gacttcgtcg cgggccactc catcggcgag 3720

atcgcggccg cacacgtggc cggagtgttc tccctggagg acgcctgcat gctcgtcgcc 3780

gcacgcggcc gcctcatgca ggcgctgccg gccggtggcg tgatgatcgc gctgcaagcg 3840

tccgaggacg aggtgctgcc gctgctcacc gaccgggtga gcatcgccgc gatcaacggc 3900

ccgcaggccg tggtcatcgc cggtgacgaa gacgcggcgg ccgcgatcgc cgagaccttc 3960

caggccgcgg gccgcaagac caagcggctg acggtcagcc acgcgttcca ctcgccccac 4020

atggacgcca tgctggagga attcctccgc gtcgcccagg tgctggacta cgccaagccc 4080

accctccccg tcgtctccct cctcaccggc accaccgcga cccccgccga actggccacc 4140

cccgcatact gggtgcgcca cgtccgggac gccgtccgtt acctcgacgg cgtacgcacc 4200

ctccaccagc ggggcgtacg caccttcctg gaactcgggc cggacgcggt gctcaccgcc 4260

atggcacagg actgcgtcga cccgcagggc gccgccttcg cccccgcgct gcgctccggc 4320

cgcccggagg cggccactgt gctcaacgcc gtcgcgcacg cccacgtccg gggtgcggag 4380

acggactggg ccgcgttctt cgccggtacg ggcgctcagc gggtcgatct gccgacgtac 4440

gccttccagc ggcagcgcta ctggatggac tcccgcaccc cggccccgga ctccgccgcg 4500

cagcgggcgc acggcggcgc cgatccggtc gaccgtgtgt tctgggacgc cgtcgagcac 4560

gaggacgtgg ccacgctcgc cgccgccctc gaactcgacc tcgacggcga acagccgctc 4620

agcgaggtcg ttccggcact gtccgcctgg cgtcgccgcc gccgcaccca gtcggaggtg 4680

gacggctggc gttaccgggt gacgtggaag ccgctgactg aggtctcgac gtctgggttg 4740

tccggttcct gggtggtgat ctcgccagct gggggtgccg atgactcggc tgtggtgagt 4800

gcgctggttg ggcgtggtgt tgacgtccgt cgggttgtgg tcgaggcggg tgtggaccgt 4860

tcggcgctgg ctgggttgct ggctgaggtt ggttcgcctt cgggtgtggt gtcgcttctc 4920

gggctggatg agtccggggg gttgttgggg actgttggtt tggtgcaggc gttgggtgat 4980

gccggggtgg gggcgccgtt gtggtgcctg actcgtggtg cggtgtctgt ggggcgttcg 5040

gatcggcttg tgtcgccggt tcaggcgcag gtgtggggtt tggggcgggt tgctgctctg 5100

gaggttccgg agtggtgggg cgggctcatc gatctgcctg aggtgctgga cgagcgggct 5160

gtgtcccgct tggtcggtgt acttgcgggt tccggtgagg atcaggtcgc ggttcgttcg 5220

tctggtgtgt tcggtcgtcg tctggtgcgt gcaccgcggg ccgagggtgc ttcggcgtgg 5280

tctccgaccg gcacggttct cgtcaccggt ggtacgggtg tgctgggtgg ccgggtggcg 5340

cgttggctgg cgggggcggg tgctgagcgt ctggtgctga ccagccgtcg tgggctggat 5400

gcgccgggtg cggttgagct ggtggaagag ctgaccaccg gctttggggt ggaggtttcg 5460

gtcgtcgcgt gtgatgcggc cgaccgtgac gccctgcgtg ccctgctgtc cgctgaggcc 5520

gggtctctga ccgctgtggt gcacacggcc ggtgttctgg acgacggcgt cctggatgct 5580

ctgaccccgg accgtatcga cagcgtcgtg cgtgcgaaag ccgtctcggc tctcaacctg 5640

catgagctga cggccgagct gggtatcgag ctgtccgact tcgtcctctt ctcctccgtc 5700

acaggtacgg tcggcgcggc cggacaggcc aactacgccg ctgcgaatgc cttcttggat 5760

gctctggccg agcagcggcg cgccgatggt ctcgcggcga cgtccatcgc gtggggtccg 5820

tgggccgagg gaggcatggc cgccgacgag gcgatggacg cacggatgcg ccgcgagggc 5880

atgcccccga tggcgcccac atccgcgatg agcgcactgg agcaggccgt tggtgcgggc 5940

gagacggcgc tgaccgttgc cgacatcgac tgggagcgtt tctcctccgt catcgccgca 6000

gtccgcccca acccgctgat cggtgacttc gtcgtcggag cggaaggcac ggccgccgcc 6060

agcggccacg gatccgtggt caccggcgcc gatgtcgccg ccaccgtctc gggccggttg 6120

gcgggcctga cccaggccga gcaggagcgg gaactgctca gcctggtccg tctgcacgtg 6180

gccgcggtac tcgggcacga cggatcggac gcggtcggtg ccgaacgggc cttcaaggaa 6240

ctcggcttcg actccctgac ctccgtcgag ctgcgcaacc gcctcggagc cgccaccgat 6300

ctccggctcc ccaccacgct cgtctacgac taccccacgt ccgccgctct cgccgagtac 6360

ctgcggggcg aactggccgg cagcgcgcag gacgccgggc cgcccctgcc cgccgtggtc 6420

ggctccgccg ccgacgacga tccgatcgtg atcgtctcga tgagctgccg cttccccggt 6480

ggcgtacgga ctccggaaga cctgtggcag ctcctcgcgg acggcacgga cacggtcgcc 6540

gccttcccgg ccgaccgcgg ctgggacctg gacggcctct acagcgccga cccggagcgt 6600

tcggggacct cgtacacgcg tgaaggcggg ttcctctacg acgccgccga cttcgacgcg 6660

gacttcttcg ggatctcgcc gcgcgaggcc ctcgccatgg acccgcagca gcgcctgctg 6720

ctcgaaaccg cctgggagac cttcgagcgc gccgggatcg acccggcgtc gctgcggggc 6780

agccaggccg gtgtcttcgt cggcaccaac ggccaggact acctctcgct ggtcacgcgc 6840

gaaggcgacg gactcgacgg actcgaagga catgtcggca ccggcaatgc ggccagtgtc 6900

gtctccggcc ggctctctta cgtcttcggt ctcgaaggcc cggcgatcac ggtcgacacg 6960

gcctgctcgt cgtcgttggt cgccctgcac ctggccgtgc aggcgctgcg ccagggcgag 7020

tgcaccttgg cgctcgccgg tggtgtgacg gtgatgtcca ctccggacgc cttcgtcgac 7080

ttcagccgtc agcgtgggct cgcggaggac ggccgtatca aggcgttcgc gtcggccgcg 7140

gacggtacgg gctggggtga gggcgtcggc atgctcctgg tggagcggct gtccgacgcc 7200

cgtaggaacg gtcacccggt cctggcggtc gtgcggggct cggcgatcaa ccaggacggc 7260

gcgagcaacg gcctgaccgc gccgaacggt ccgtcccagc agcgcgtcat ccgccaggcg 7320

ctggccggtg cggggctgtc ggccgccgac gtggacgcgg tggaggcgca cggtacgggc 7380

acccggctcg gtgacccgat cgaggcgcag gcgctgctcg ccacgtacgg ccaaggccgc 7440

ccggcggacc ggccgttgtg gctgggctcc gtgaagtcga acatcggtca cacgcaggcc 7500

gccgcgggcg tggcgggcgt gatgaagatg gtcatggcga tgcggcacgg tgtgctcccg 7560

cgcacgctgc acgtggacgg gccgaccccg cacgtcgact ggtcggcggg cgacgtcgcc 7620

ctgctgaccg agcagcggga gtggccggcg accggccacc cgcggcgggc aggtgtgtcc 7680

tcgttcggcc tgagcggtac gaacgcccac accatcatcg aagaagcccc ggccgacgac 7740

gacgccgagc ccacgaccgg cgcggggacg gccccgtccg ttctgccgct gctcatctct 7800

gccaagagcg acgccggcct gcgcgcacag tcggagcagc tggcgaccca tctggtcgga 7860

aacccggacg tccccatcgg ggacatcgcc tactccctca cgaccggacg ctccgggctg 7920

gagacgcgag cgatcctggt cggcgacgcc gacaaccgca cagggctcgc ggccgcgctg 7980

cgaagcctcg ctgccggcga gcaggctccg ggcctggtcc agggcacggt gaccgagggc 8040

gggctggcgt tcctgttcac ggggcagggg agccagcggc tggggatggg ccgtgagctg 8100

tacgagacgt atccggtgtt cgcggatgcg ctcgacgcgg tgtgcgcgcg gatggatctc 8160

gaagtcccgc tgagggacgt gctgttcggg gcgtatgcgg gtctgctgga tgagaccgcg 8220

tatacgcagc ctgcgttgtt cgcggttgag gtggcgttgt tccggctggt ggagagctgg 8280

ggtctgaggc cggacttcgt ggcgggtcat tcgattggtg agatcgctgc tgcgcatgtg 8340

gcgggggttc tgtccctgga tgacgcctgt gctctggtgg aggcgcgtgg gcggttgatg 8400

ggtgcgctgc ctggtggtgg cgtgatgatc gcggtccagg cgcctgaggc tgaagtcctg 8460

ccgctgctga ccgagcgcgt gagcattgcc gcgatcaatg gtccgcagtc ggtcgtgatc 8520

gcgggtgacg aggccgacgc ggtggcgatc gtggagtcgt tcacggggcg taagtccaag 8580

cggctcacgg tcagccacgc gttccattcg ccgcacatgg acggcatgtt ggaggacttc 8640

cgggccgtgg cggaagggct gtcgtacgag gccccgcgca tccctgtggt ttccaacctc 8700

accggggccc tggtctcgga tgagatgggg tcggctgagt tctgggtgcg tcatgtccgc 8760

gaggcggttc gcttcctgga cgggatgcgt gttctggagg ccgccggggt tacgacgtac 8820

gtcgagcttg gcccgggggg tgtgctgtcg gcgctggcgc aggagtgtgt cagtggggac 8880

ggtgctgctt tcgtgccggt gctgcgttct ggccgtcccg aggccgagac cgcggtcacc 8940

gcgttggccc aggcacatgt gcggggtgtg gacgtcgact gggccgcgtt cttctccggg 9000

accggcgtcc agcgggtcga cctgcccacc tacgccttcc agaggcagcg gttctggccc 9060

gcgatgacgg cggagagtgc gccggtcggc gggacggtcg acgcggtgga cgcccacttc 9120

tgggatgtca tcgagcagga ggacgtcgag tcccttgctg agttgctcgg tctcgacgac 9180

gcgagcgcgt gggggagtgt ggtccccgcg ctctcggcct ggcgtcggca gggccaacag 9240

caggcccagg tcgacggatg gcgctaccgg gcgagctgga agccggtgac ggctgcggtg 9300

tcgtccggcg tggtgagcgg gacatgggtt gtcgccgtac ctgccggatc tgcgggggac 9360

gacgcgcggg tcgaggccgt gaccaacggg ctggctgggc gtggcgttga cgtccgtcgg 9420

gttgtggtcg aggcgggtgt ggaccgggcc gcgctggctg ggttgctggc tggtgaggga 9480

tctctcgctg gtgtggtgtc gcttctcggg ctggatgagt ccgggggggct ggcggctact 9540

gctggtttgg tgcaggcgtt gggtgatgcc ggggtgtcgg cgccgttgtg gtgcctgacc 9600

cgcggggctg tttccgtcgg tcgttcggat cggcttgtgt cgccggttca ggcgcaggtg 9660

tggggtctgg ggcgggttgc tgctctggag gttcccgagc gttggggcgg gctggttgac 9720

cttccggaag tgctggatga gcgggctgtg tcccgcttga tcggtgtact tgcgggttcc 9780

ggtgaggatc aggttgcggt tcgttcgtct ggtgtcttcg gtcgtcgtct ggtgcgtgca 9840

ccgcgggccg agggtgctgc gtcgtggact ccgaccggca cggttctcgt caccggtggc 9900

acgggtgtgc tgggtggccg ggtggcgcgt tggctggcgg gggcgggtgc tgagcgtctg 9960

gtgctgacca gccgtcgtgg gctggatgcg ccgggtacgg ctgaactggt cgaggagctg 10020

accagctccg gggtggaggt gtcggtcgtc gcgtgtgacg cggccgaccg tgacgccctg 10080

cgcgccctgc tctcctctga ggccgggtct ctgaccgctg tgatccacac ggccggtgtc 10140

ctggacgacg gtgtcctgga tgctctgacg ccggaccgta tcgatggtgt cgtgcgtgcg 10200

aaggccgtct cggctctcaa cctgcacgaa ctgacggccg agctgggcat cgagctgtcc 10260

gccttcgtcc tgttctcgtc catgagcggc acggtgggca cggcgggtca ggccaactac 10320

gcggctgcca atgcctacct ggatgctctg gccgagcagc gccgggcgga cggtctcgcg 10380

gcgacgtcca tcgcttgggg tccgtgggcg gagggtggca tggccgccga tgcggcgctc 10440

gaagcccgta tgcgccgaga cggggtgcct ccgatgcccg cggatccggc gatccgcgct 10500

ctccggcagg ccgttgcagg cgacgacgcc gtgcttaccg ttgccgatgt cgaatgggac 10560

cggttcctcc cgggcttcgt cgccgcacgg cacagcgagc tgttcagcga gctgcgtgac 10620

gtccgtgatg cccgcgcggc acaggatcgg gcgcaggccg ccgttgccgc cgaccgtccg 10680

gactcccttt ccgggcggct gtccgcccag gcgccggccg agcaggagcg agagctgctg 10740

gacctggtcc gtacgcaggt cgccgccgtg ctcgggcacg ccggagtgga aaacgtgggc 10800

gcggggcggg cgttcaagga gcttggcttc gactcgctca tggccgtcga gctgcgcaac 10860

cgcatcggct cggccaccga gcttcggctc ccggccacct tgatctacga ccaccccacg 10920

tccgccgccc tcgcggagtt cctgcggggt gagctggtcg gcaccgtgcg ggtcgccgac 10980

aaggtgctgc ccgccgtggt ctccgccgac gaggatccga tcgcgatcgt ctcgatgagc 11040

tgccgcttcc ccggtggcgt acggactccg gaagacctgt ggcggctcct cgtggacggc 11100

acggacgccg tcggcgcgtt cccggccgac cgcggctggg acctggacag gctctacagc 11160

cccgacccgg accagccggg cacctcgtac acccgcgaag gcgggttctt cgacggggcc 11220

gcggacttcg atcccgggtt cttcgggatc tcgccgcgcg aggcgctcgc catggacccg 11280

cagcagcgac tgctgctcga aacctcctgg gaggcgatcg agcgggcggg catcgacccg 11340

tcgtcgctgc gcggcagcca ggccggtgtc ttcgtcggca ccaacggcca ggactacctc 11400

tcctcatca cccgtgaatc ggagggcctg gaaggtcact tgggcacggg taacgcgggc 11460

agcgtcatgt ccggccgcgt ctcctacgtg ctcggcctgg agggtccggc ggtcacggtc 11520

gacacggcgt gctcgtcctc gctggtcgcc ctgcactggg cgatccaggc cctgcgtcag 11580

ggcgagtgca gcatggctct ggccggcggc gtgaccgtca tgtcgacgcc cgagaacttc 11640

gtcgacttca gccgtcagcg cgggctcgcg gaggacgggc gcatcaaggc gttcgcgtcg 11700

gccgcggacg gtacgggctg gggtgagggt gtcggcatgc tcctggtgga gcggctgtcg 11760

gatgcccggc gcaacgggca tccggttctg gcggtagtac gtggttcggc tgtcaatcag 11820

gacggtgcga gcaatggtct gacggctccg aatggtcctt cgcagcagcg ggtgatccgt 11880

gcggcgctgg cgagtgcagg tctgtcggcc gctgatgtgg atgtggtgga ggcgcacggt 11940

acggggacga agctgggtga cccgatcgag gcgcaggcgc tgctggcgac gtacgggcag 12000

gaccggcccg cgggccgtcc gctgtggctg ggttccatca agtcgaacat cggtcatacg 12060

caggccgccg ccggtgtcgc gggcatcatc aagatggtcc tcgccatgca gcacggcgtg 12120

ctgccgcaga cgctgcacgt cgacgagccg accccgcacg tcgactggtc ggcgggcgag 12180

gtcaccctgc tgaccgagca gacggcctgg ccgacggtgg accggccgag gcgagcggga 12240

gtgtcgtcct tcggcatcag cggcaccaac gcccacacca tcatcgaaca ggccccggcg 12300

gtcgagcagt tggcggacgg tgacgcgact cccgccactc cggccctcgc gctcccgctg 12360

ccgtacgtcc tctccgcgaa gagccccgag gccctgcgcg cccaggcgtc cgtactgcgc 12420

acgcacctgg aggccacgga ccacaacggg cccggttccg acgacctggc cttctcgctc 12480

gccacggcac gtgcgcacct cgaacaccgc gcagtcctga ccgccgacga cccacaggaa 12540

ttccgggagg cactcgcacg cctcgccgac ggtgatccct caccgaggat caccaccggg 12600

gcggtgagcg acggtcgtac ggcgttcctg ttcacgggcc aggggagtca gcggctcggg 12660

atgggccgtg agctgtacga ggcgtatccg gtgttcgcgg acgcgcttga cgcggtctgc 12720

gcgcatgtgg acgcgcacct cgaagtgccc ctgaaggacg tcctgttcgg ggcggatgcg 12780

ggtctgctgg accagacggc ttacacgcag cccgcgttgt tcgcggtcga ggtggcgttg 12840

ttccggctgg tggagagctg gggtgtgaag ccggacttcg tggccggtca ttcgatcggt 12900

gagatcgcgg ccgcgcatgt ggcgggcgtc ttctcgctcc aggacgccag tgaactggtc 12960

ttcgctcgtg ggcggttgat gcaggcgctg ccgaccggtg gcgtgatgat cgcggtccag 13020

gcgtcggagg acgaggtcct gccgctgctg accgaccggg tgagcattgc cgcgatcaac 13080

ggcccccagt cggtcgtcat cgcgggcgac gaggccgacg cggtggccat cgccgagtcc 13140

ttcacggacc gcaagtccaa gcgcctcacg gtgagccacg cgttccactc gccgcacatg 13200

gacggcatgc tcgacgcctt ccgtgagatc gccgagggcc tctcctacga accttcgcgc 13260

atcccggtcg tctcgaacct caccggcgct ctcgtctccg atgagatggg ctcggccgag 13320

ttctgggtgc ggcacgtccg cgaggccgtc cgtttcctcg atggcatccg cacgctggaa 13380

gccgcgggcg tcaccaagta cgtcgaactc ggccccgacg gcgtgctgtc ggcgatggcc 13440

caggactgcg tgagtggcga gggctccgtc ttcatccccg tgctccgcaa ggcgcgcccc 13500

gaggccgaga gcgtcacgac cgccctcgcc tcggcccacg tccacggcat ccccgtcgac 13560

tggcaggcgt acttcgccgg gaccggcgcc cagcgcgtcg acctccccac ctacgccttc 13620

cagcgccagc gctactggcc cagcgctgcc gcgttcgtca ccggcgatcc gacggcgatc 13680

gggctcgggg atgccgggca cccgttgctg ggtgcggcgg tggcgctcgc cgactccgag 13740

ggcgtgctct tcaccggccg cctgtcgctc gacacccacc cctggctcgc cgaccacacc 13800

atcctcggca gcgtcctgct gccgggcacg gccttcgtcg acctggcgat ccgggccggc 13860

gatcaggtcg gatgcgatgt ggtcgaggag ctgaccctcg aagcgcccct cgtcgtcccc 13920

cagcggggcg gtgtgcagct ccagctcgtc gtcgaggcgc cgagcgggcc cgggcagcgg 13980

ccgttcagcg tgcactcccg gcggcaggac gcctacgcgg aggagccgtg gatgcggcac 14040

gcctccggag tgctgacttc cggcgtttcc cgccgcgaac tgtccgtgga aggcggggag 14100

ttcgaggcgc tggccgtctg gccgccgacc ggagccgtac ccgtggacgt acgaggtctg 14160

tacgaggagc tcgccgaggc cggtgtggcc tacgggccgc tgttccaggg gctcaaggcg 14220

gcgtggcggc gggacggtga actgttcacc gaggtggcgc tcccgggtga agcccggcgt 14280

gaggcggcac ggttcggtct gcacccggct ctgctggacg ccggtctgca cgccatcggc 14340

cacggcgagg gaccggaacc ggcaatgacc ggcgcgctgt tgcccttctc ctgggcagga 14400

gtctcgctgt acgcggcggg cgcctcctca ctcaggatgc ggctgacccc gcacacaccc 14460

gacgacgccc acaccatggc gttgctcgtg gcggatgaga ccggacgtcc ggtggcggcc 14520

gtggagtcgc tgatcctgcg taccgcgtcg gccgaccagg tgcgcgcggc cgacggaggt 14580

cacctcgact ccctcttcaa ggtggagtgg ctgcccgtgg cgggcggagc cacgccgcac 14640

ggcgactcca ccggacggcg atgggccgtc ctgggccgcg acggactcgg cctgccggcc 14700

accggcgtgc aggggcaggt ggccgagtac gacgatgcct ccgcgctcgg tgcggcgctc 14760

gcggccggcg aaccggtgcc ggacgccgtg ttcgtccacc ctggggctct tccggggcag 14820

gacacggaca ccacggcggc ctccgtacac gccgccgtga cggacgcgct gtccttcgta 14880

caggaatggc tggcggacga gcggttcgcc gccacgcgcc tggtgtggct gacatccggc 14940

gcggtggcgg acgagcccgg cgcgggcgtc cgggacctgg cgggcagcgc cgtacgcggc 15000

ctgctgcgct cggcgcagtc cgagaacccc ggccagctgc tgatgctcga cctcgaccag 15060

gacccggcct cgctcgcggc gctgcccgcc gcgctggccg cgggtgagcc ggaactggcg 15120

atacgacgcg gagaactccg taccccgcgc ctgacgcgcg tcccctcggc ggacgccgcg 15180

gcagagccgc tcggcacact cggcgacccg tccggcacgg tactcgtgac cggagccacc 15240

ggaaccctgg gcggactctt cgcccgccat ctggtgacgg cgtacggggt gcggcgactg 15300

ctgctcacca gccgtcgcgg ccccgaggcc gaaggtgcgg ccgaactggt cgccgaactg 15360

gagcagttgg gggcgcacgt cgaactcgtc gcctgcgacg ccgccgaccg ctccgcgctc 15420

gccgcgctcc tcggagccgt accgtccgag cacccgctga cggccgtggt gcacacggca 15480

ggcgtactgg acgacggcat cctctcctcg ctcacccccg agcgcgtggc cgccgtactg 15540

cgtccgaagg tggacgccgc ctggaacctg cacgagctga cgcgggaact cggcctctcg 15600

gcgttcgtgc tcttctcggg cgccgccgcc gcgttcggcg cggccgggca ggggaactac 15660

gccgccgcca acagcttcct ggaagccctg gcggagcagc gccgcgccga aggcctgccc 15720

gccacctcac tcgcgtgggg cctgtgggct ccgcagacgg gcggcatggc ccagcagctg 15780

gacgaggtcg acctgcggcg catcgccagg gacggcgtcg gcgggctctc cggtgacgag 15840

ggcctcggcc tcttcgacac cgcgatgacg gtcgacgcgg cggtcctgct gcccatgcgg 15900

ctcgacctcg cggtggcgcg ggcgcaggcc gtctccacgg gcgagacacc ggcgctgctg 15960

cgcgccctca tacgggtgcc cgccgcggcgc gcggtcgagc agcgtacggc ggcggacggg 16020

gcctcgcccc tggcggccag gctgtccgcc ctgccggacg cggaacgcga ggacatgctg 16080

ctggacctgg tgtgcgggcg ggtggccgag gtcctcggcc acaccgacgc ccgcgcggtc 16140

gacgcggacc gcgcgttcaa ggaactcgga ttcgactccc tcacggccgt cgagctgcgc 16200

aacgtcctga aggccgcgac cggcctcagg ctctcaccga ccctcgtctt cgactatccg 16260

accccggtgg cgctggcccg gcacctgctc gccgagctgg cgggaaccgc cgatgaccag 16320

gacgccgtac gcggccggaa ggcacccgca cggcccgcca cggccgcggt cacctccgtg 16380

accggcgaag acccgatcgt catcgtcggc atgggctgcc gcttccccgg cggcgtacgg 16440

tcgccggagg acctgtggca gctcgtcgcc accggcggcg acggcatcac cggcttcccg 16500

tccgaccgcg gctggaacgt cgaggccctc taccaccccg acccggacca cgcaggcacc 16560

tcgtacaccc gcgaaggcgg cttcctgcac gacgccgccg acttcgatcc cgggttcttc 16620

gggatctcgc cgcgcgaggc cctcgccatg gacccgcagc agcgcctgct gctggaaacc 16680

tcgtgggagg cgttcgagcg ggccggaatc gacccggcga cgctgcgcgg aagccgtacg 16740

ggcgtcttcg ccggtgtcat gtaccacgac tacgtgaccg gcatcggcga cggcggcagc 16800

gccgtcgaac tgcccgaggg ggtcgagggc tacctcggca ccggcaacgc cggcagcatc 16860

gcctccggcc ggatcgccta caccttcggc ctcgaaggcc cggcggtcac cgtcgacacg 16920

gcctgctcct cgtcgctcgt cgccctgcac tgggcgatcc aggcgctgcg cagcggcgag 16980

tgcacgatgg cactggccgg cggtgtcgcc gtcatggcca ccccgagac cttcgtcgac 17040

ttcagccgcc agcgcggcct ctcggccgac ggtcgctgca agtccttcgc cgcggcggcg 17100

gacggtacgg gctgggccga aggcgcgggc atgctcctgg tggagcgcct ctccgacgcc 17160

gaacgcaacg ggcacccggt cctggccgtg gtccgcggct cggcgatcaa ccaggacggc 17220

gcgagcaacg gcctgaccgc accgaacggt ccgtcccagc agcgcgtcat ccgcgaggcg 17280

ctggccagtg ccgacctgtc ggccgccgac atcgacgcgg tcgaggccca cggcacgggc 17340

acccggctcg gcgacccgat cgaggcgcag gcactcctgg ccacgtacgg ccgtgagcgc 17400

gaggcgggcc gcccgctgtg gctcggctcg atcaagtcga acatcggtca cacgcaggcg 17460

gcggccggtg tcgcgggcat catcaagatg gtcatggcga tgcggcacgg cgtactgccg 17520

cagaccttgc acgtcgacga gccgtcaccg caggtcgact gggaggccgg tgaggtctcc 17580

ctgctgaccg gggcgatgcc ctggccgcag acgggccgtc cgcgccgtgc gggcgtgtcg 17640

tcattcggca tcagcggcac caacgcccac acgatcatcg agcagccgcc gacccgtgag 17700

gtgacgccga cggttccggt ggctccggtg gttccgacgg ttccgacggt tccggtggtg 17760

ccgtgggtgc tctcgggcaa gggcgaggag gcgctgcgag cgcaggcacg tcagctccag 17820

tcgtacgtgc tccgcgcacc ggaactgcgt ccggtcgaca tcgccggctc gctggcggtg 17880

ggccgggcgt ccttcgagga ccgcgcggcg gtggtcgccg ccgaccgcga ggggcttctg 17940

gccgcccttg cggcgctggc ggacggcggc tcggcgacgg gggctgtgga gggttccgcg 18000

gtgggcggga agctggcgtt cctgttcacg gggcagggga gccagcggct ggggatgggg 18060

cgcgagctgt acgaggcgta tccggtgttc gcggaggcgt tggatgcggt gtgtgctcgt 18120

cttgaactgc ctttgaagga tgtgttgttc ggggcggatg cgggtctgct ggatgagacc 18180

gcgtatacgc agcctgcgtt gttcgccgtt gaggtggcgt tgttccggct ggtggagagc 18240

tggggtctga ggccggactt cgtggcgggt cattcgattg gtgagattgc tgccgcccat 18300

gtggcggggg tgttctcgct ggatgacgcc tgtgctctgg tggaggcgcg tgggcggttg 18360

atgggtgcgc tgcctgcggg tggcgtgatg atcgcggtgc aggcgtcgga ggacgaggtc 18420

ctgccgttgt tgaccgaccg ggtgagcatt gccgcgatca acggtcctcg gtcggtggtg 18480

atcgcgggtg acgaggccga cgcggtggcg atcgtggagt cgttcacggg gcgtaagtcg 18540

aagcggctta cggtgagtca cgcgttccat tcgccgcaca tggacggcat gttggaggac 18600

ttccgggccg tggcggaggg cctgtcgtac gaggccccgc gcatccccgt cgtctccaac 18660

ctcaccggca ctctcgtcac cgacgagatg ggctcggctg agttctgggt gcgtcatgtc 18720

cgtgaggcgg ttcgcttcct ggacggtatt cgggctttgg aggctgctgg ggttacgacg 18780

tatgtcgagc ttggccctgg gggtgtgctg tcggcgctgg cgcaggagtg tgtcagtggg 18840

gacggtgctg ctttcgtgcc ggtgctgcgt tctggacgtt ccgaggccga gactgcggtg 18900

accgcgttgg cccaggcgca tgtgcggggt gtgaacgtcg actgggccgc attcttcgcc 18960

gggaccggcg ctgagcgggt cgacctgccg acgtacgcct tccagcggca gcgctactgg 19020

ctgcacatcc cccgcgtcgc gcagagcggg gtcgccgacg aggtggacgc ccggttctgg 19080

gatgccgtgg agcgtgagga tctggagtcg ctcgcctcca ccctggaggt cgacgacgag 19140

agcgcgtgga gcagcgtctt gcctgcgctg tcggcgtggc gtcgggagcg gcgtgcccag 19200

tccgaggtgg acggttggcg ttaccgggtg tcgtggaagc cgctggctga ggtctcggcg 19260

tcggggttgt ccggttcctg ggtggtgatc tcgcctgctg ggagtgtgga cgactcggct 19320

gtggtgagtg cgctggttgg gcgtggtgct gaggtccgtc gggttgtggt cgaggcgggt 19380

gtggaccgtt cggcgctggc tgggttgctg gccgatgcgg gttctgccgc gggtgtggtg 19440

tcgcttctcg ggctggatga gtctgagggg ttgttgggga ctgttggttt ggtgcaggcg 19500

ttgggtgatg ccggggtgga ggcgccgttg tggtgcctga ctcgtggtgc ggtctccgtc 19560

ggtcgttcgg atcggctggt gtcgccggtt caggctcagg tgtggggtct ggggcgggtt 19620

gccgccctgg aggttccgga gcgttggggc gggctggttg acctgccgga agtgctggat 19680

gagcgggctg tggcccgctt ggtcggtgta cttgcgggtt ccggcgaaga tcaggtcgcg 19740

gttcgttcgt ctggtgtgtt cggtcgtcgt ctggtgcgtg caccgcgggc cgagggtgct 19800

tcggcgtgga caccgaccgg cactgttctt gtcaccggtg gtacgggtgt gctgggtggc 19860

cgggtggcgc gttggctggc gggggcgggc gctgagcgtc tggtgctgac cagtcgtcgt 19920

ggtccggatg ctccgggtgc ggctgagctg gtggaggagc tgaccaccgg cttcggggtg 19980

gaggtttcgg tcgtcgcgtg tgacgcggcc gaccgtgacg ccctgcgcac cctgctctcc 20040

gccgaggccg ggactctgac cgctgtgatc cacacggccg gtgttctgga cgacggcgtc 20100

ctcgacgcgc tcaccccgga ccgtatcgac agcgttctgc gtgccaaggc tgtctcggcg 20160

ttcaacctgc acgagctgac ggccgagctg gggatcgagc tgtccgcctt cgtgctgttc 20220

tcgtcgatga gtggcacggt gggtgcggcc ggtcaggcca actacgccgc tgccaacgcc 20280

tacctggatg ctctggccga gcagcggcgc gccgatggtc tcgcggcgac ctcgctcgct 20340

tggggtccgt gggccgaggg cggcatggcc ggcgacgacg cgatggacgc acggatgcgc 20400

cgcgaggggc tgccccccgat ggcgccggac gcggcactga ccctgctgcg tcagagcgtg 20460

gggtccgccg atgcggcgct gatggtggtc gacgtggagt ggcagcggtt cgcccctgcc 20520

ctgaccgtcg tgcgccccag caacctcctc gccgagttgc ccgaggctcg ccccgccgga 20580

acggattccc gtacgggtgg cgcaacgtcc tccgaggggg ccggctcgtt cgccgagcgg 20640

ttggccgccc tgggtggggc cgagcaggac aaggagctgc tgaacctggt ccgtacgcat 20700

atcgccgccg tactcggaca tggcggctcg gaggccgtgg gtgccgaacg ggccttcaag 20760

gaactcggct tcgactccct gaccgccgtc gagctgcgca acaggctcgg tgccgcgacc 20820

ggtgtacgtc tcccggccac gctgatcttc gactacccga ccgccacggc tctcgccgcc 20880

tacctgcggg gcgagttgct cggtacgcag gtcgtggtgt ccggtccggt gtccaacggc 20940

gtcgtcgtgg acgacgatcc gatcgcgatc gtcgcgatga gctgccgctt ccccggtggc 21000

gtacggacgc cggaagacct gtggcggctg ctgtcgaccg gcggtgacgc catcggtgag 21060

ttccccgccg atcgcggctg ggatctgagt cggctctaca gccccgaccc cgacaagcag 21120

ggcaccttct atgcccgcgc gggcggtttc ctctacgacg ccgccgactt cgacgcggac 21180

ttcttcggga tctcgccgcg cgaggccctc gccatggacc cccagcagcg actgctcctg 21240

gagacgtcct gggaggcctt cgagcgggcg ggcatcgacc cgtcgtcgct gcgcggcagc 21300

caggccggtg tcttcgtcgg caccaacggc caggactacg gagcgatgct ccagaccatc 21360

ccggacggca tcgagggctt cctcggtacg ggcaacgcgg cgagcgtcgt ctccggccgg 21420

ctgtcctacg ccttcgggct cgaaggtccg gccgtcacgg tggacaccgc ctgctctgcc 21480

tcgctggtcg cccttcactg ggcggtccag gcgctgcgca gcggcgagtg ctcgctcgca 21540

ctggccggtg gcgtgaccgt catgtcctcg cccggtgcct acatcgactt cagccgtcag 21600

cgtgggctcg cggaggacgg tcgtatcaag gcattcgcgg cagccgcgga cggtacgggc 21660

tggggcgagg gcgtcggcat gctcctcgtg gagcggctct ccgacgcccg caggaacggt 21720

cacccggtcc tggccctggt ccggggctcg gccatcaacc aggacggcgc gagcaacggc 21780

ctgaccgcgc cgaacggccc ctcgcagcag cgtgtgatcc gccaggccct ggccaacgcg 21840

ggcttgtccg ccgcggaggt ggacgcggtc gaggcgcacg gcaccggcac gaggctcggc 21900

gacccgatcg aggtgcaggc actcctggcc acgtacggcc gtgagcgcga ggccgaccag 21960

cccctgtggc tcggctcgat caagtcgaac atcggccaca cgcaggcggc cgccggtgtc 22020

gcgggagtca tcaagatggt cctcgccatg gagcacgggg tgctgccgca gaccctgcac 22080

gtggacgagc cgactccgca cgtggactgg tcggcaggcg atgtcgccct gctgaccgac 22140

gccgtggagt ggcccgagac cggtcgcccg cgtcgagcgg gtgtgtcgtc gttcggcttc 22200

agcgggacga acgctcacac ggttctggaa caggcaccga agcccgagga gcctgaggag 22260

tctcagcagc ctgaggagac gaacgcgccc gcccgaccgc atcagtccgg agtcatgccg 22320

tggacgctct cggcgaagag cgaggcggcg ctgcgggtcc aggccgagcg gctgcggacg 22380

cgcatcgctt ccgacccgct gctccagccc gtcgacgtgg cctactcact cgcgacatcg 22440

agggccgccc ttgagcggcg cgccgtggtc gtcgcgacgg aacgtgacga gttcctggcc 22500

ggactcaagg cgctggcctc cgggcagcct gctccgggcc tggtgcaggg cagggtgacc 22560

gagggcgggc tggcgttcct gttcacgggg caggggagcc agcgactggg gatgggccgg 22620

gagctgtacg agacgtatcc cgtcttcgcg gatgcgctcg acgcggtgtg tgtgcgtctt 22680

gaactgccct tgatggatgt gctgttcgga accgagcgcg acgcgctgga cgagaccggg 22740

tacacccagc cggctctctt cgcggtcgag gtggcgttgt tccggctggt ggagtcgtgg 22800

ggtgtgaggc cggacttcct ggccgggcac tcgatcggtg agatcgcggc cgcgcatgtg 22860

gcgggagtgt tctcgctgga tgacgcctgc gctctggtgg aggcgcgtgg gcggttgatg 22920

caggcgctgc cgaccggcgg cgtgatgatc gccgtccagg cgtctgaggc cgaggtcctg 22980

ccgctgctga ccgagcgcgt gagtatcgcc gcgatcaatg gtccgcagtc ggtcgtgatc 23040

gcgggtgacg aagccgatgc ggtggccctc gtggagtcct tcacgggccg caagtccaag 23100

cggctcacgg tcagtcacgc cttccactcg ccgcacatgg acggcatgct cgccgacttc 23160

cgcaaggtgg cggaggggtt gtcgtacgag gccccgcgta tcccggtcgt ttcgaacctc 23220

acgggggccc tggtcaccga cgagatgggc tcggccgact tctgggtgcg gcacgtccgc 23280

gaggccgtcc gcttcctgga cggcacccgc acgctggaag ccctgggcgt cacgacgtac 23340

gtcgaactcg gccccgacgg ggtcctgtcg gcgatggccc aggagtgtgt gaccggcgag 23400

gactccgtct tcgtgccggt cctgcgctcg ggtcgtcccg aggccgagag cgtcaccacg 23460

gccctcgccc aggtacacgt ccgcgggatc gccgtcgact ggcaggcgta cttcgccggg 23520

accggcgccc agcgcgtcga cctcccgacc tacgccttcc agcgccggcg ctactggttg 23580

gaagaggctc ccgccacggc ggccgtcgag cccctgaccg gctcgctcgg ggccgtggac 23640

gcgcagttct gggcggccgt cgacaacgcg gatctctccg cgctcaccgc caccctggac 23700

atcgacgtcg acgccgacca gccactgagc gccctgctgc ccgcactgtc cgcctggcgg 23760

cggcagcgtc aggagcagtc ggtcgtcgac ggctggcgct acacggtcac atggaagccg 23820

atggccgatc cggccgtcgc acggccgacc gggacctggc tcgtcgtgac ccccgccacc 23880

agccttgtcg acctgcccgc ggtctccgcc gcgttggcag cgcagggagt ggacgtacgg 23940

gaagtcgccc tggaggcggc cgagttggat cgcgacggcg tggcgggccg gatgcgtgag 24000

gcgctcgcgg gcgaccgggc cgacggggtg ctgtccctgc tggcgctcgc cgaacacccg 24060

cacccggccc atccggcggc gcccaccggg ctgctcctga ccgggacgct cgtacaggca 24120

ctcggtgacg ccggagtcga cgccccgctg tggtgcctca ccaccggcgc cgtggcgacc 24180

gcaccctccg acctgatcgg gagcgcggcg caggcgcagg tctggggcct cggccgggtc 24240

gtcgccctgg aacaccccga gcgctggggc gggctcgtgg acctgcccgt accggcggac 24300

gagcgggcac tcgaccggct gctcgccgtc ctcgcgggcg ccggggacga ggaccagatc 24360

gccgtacggt ccgcgggcct cctcgcccgc cgcatcgggc acgccgcgcc tcccgccgcc 24420

gggcagcacg ccgacagcgg gacatcgggc gccggcgctg cggccggctc cgcctggcgg 24480

ccgcggcggca ccgtcctggt caccggaggc acgggcgcgc tcggcgggca cgtcgcccgc 24540

tggctcgcgg cacacggcgc ggaacacctg gtgctgctca gcaggagggg cccgcaggcg 24600

cccggcgccg atgccctggt cgccgagatc gccgcgctgg gtgccggggc cacggccgtc 24660

gcctgtgacg tgaccgaccg gaccgccgtg tcggagctgc tcgccgggct cgccgacggc 24720

acgtacggtc ccggcctcac cgccgtcttc cacacggcgg gcgccgggca gttcgcgccg 24780

ctcgacggga ccggccccgg cgaggtcgcc gaggtcgtcg ccgccaaggt cgcgggcgcc 24840

gcccacctcg acgagctgct cggggacacg gaactggacg ccttcgtcct cttctcctcc 24900

atcgccggcg tctggggcag cggcggccag agcgcctacg cggcggccaa tgcccacctg 24960

gacgccctgg cccagcagcg ccgggcccgc ggactgacgg ccacgtccgt ggcctggggc 25020

ccgtggggcg agggcggcct ggtcgccgac gacgaagcgg ccgaacaact gcgccgccgc 25080

ggcctgcccg tcatggcgcc ggagctgtcg atcgccgccc tccagcaggc gctggacggg 25140

gacgagacgg cggtgacggt ggccgatgtc gactgggacc tgttcgtgcc ggccttcacc 25200

gccgcccggc cgcgtccgct gatcaccgac ctccccgagg tgcgccgcgc tctggcggca 25260

gagcaggacg gagccgccac cgcggccggg gaagcggccg gcctcgaagc cgagctgcgg 25320

gggatgagcg gaaccgaggc ggagggcgtc gtcctgaacc tggtccgtac gcaggtcgcc 25380

gtcgttctcg gacacggggg agcgacggcg gtcgaggcgg cccgcgcctt caaggaactg 25440

ggcttcgact cgctcaccgc ggtcgagctg cgcaaccgcc tcagcaccgc caccggactg 25500

cggctgcccg cgagcctggt cttcgactac ccgaccccgg ccgcactggc cgcgcacatc 25560

cgggcggaac tcctcggcga ggacaccacg cccgaactgc ccgccctcgc ggagatcgac 25620

aagctggaat tcctcctctc gtcggttccc gaggacacca ccgaacgcgc ccgcgtcacc 25680

gcacggctcg aatcgctcct gtcgaactgg aacagggcag aacgagcggt catcggagag 25740

gacgaagaaa tatccatcga atcggcatcc gccgacgacc tcttcgacat catcaacaac 25800

gaattcggaa aatcctgacc tgatgaccga tccgatgacc gatccgaatt ccgatccaat 25860

gtccgtatgc attccgcaat tccccaggag gtgacgttcc agtggccagc gcgaacgaag 25920

aaaagcttct cgaaaacctg aagtggatga ccaatgagct gcggcgggcc cgccgtcgcc 25980

tccatgaggt cgaggcggac gcccaggaac cgatcgcgat cgtcgcgatg agctgccggt 26040

tccccaacgg ggtgggatcc ccggaggatt tgtggcgcct ggtcgacgag ggcggcgacg 26100

ccatcaccgg attccccgcc gaccgcggct gggacatcga gtcgctcgcc gatccggacc 26160

ccgaccgcaa gggcaccttc tacaacaccg gcggcggatt cctcgacggg gccaccgcat 26220

tcgatcccgg atttttcggc atatcgcccc gcgaagcgct cgccatggac ccgcagcagc 26280

gccagctcct ggagacctcg tgggaggtat tcgagcgcgc gggcatcgac cccgcggccg 26340

tacgcggcag ccgcaccggc gtctacgtcg gcgcgggcgc gatggggtac ggagccgacc 26400

tcaaggaagc gccggaaggg ctggagggac tgctgctgac cggcggcgcc accagcgtcc 26460

tgtcgggacg ggtcagctac gtgttcggac tggagggccc cgccgccacc gtcgacacgg 26520

cctgctcctc ctcgctcgtc gccctgcacc tcgccaccca ggccctgcgt cagcgcgagt 26580

gctcgctcgc gctggtcggc ggcgtgtgcg tgatgcccag ccccgatgtg ttcgtcgagt 26640

tcagccgcca gcgcggcctg tcgcccgacg gccgctgcaa gtccttcgcc gcgtccgccg 26700

acggcaccgg ctggtccgaa ggcgtcggtg tcctcctggt ggagcgcctc tccgacgccc 26760

gtaggaatgg tcatccggtc ctcgcggtgg tgcgtggctc ggccgtcaat caggacggcg 26820

ccagcaacgg cctgaccgcc cccaacgggc ccgcccagca gcgcgtcata cgccaggccc 26880

tggagaacgc ccggctgtcg gcggccgagg tcgacgtcgt cgaggcccac ggcacgggga 26940

ccacgctcgg cgaccccatc gaggcccagg cactcctcgc gacctacggg caggaccgcc 27000

ccgagggccg ccccctgcgc ctggggtccc tcaagtccaa catcggtcac acgcaggccg 27060

ccgcgggtgt cgcgggcatc atcaagatgg tcatggcgat gcggcacggc gtactgccgc 27120

agaccctcca cgtcgacgag ccgaccccga acgtcgactg gaccgcgggc gccgtttccc 27180

tgctcaccga gccgatgccc tggcccgaga ccggcgcgcc ccgccgcgcg gccgtctccg 27240

cgttcggcgt gagcggcacc aacgcgcaca ccatcatcga acaggccccc gagccggacg 27300

ccgagtccgt gtccgtgtcc ggctccgcgc ccgcggcggc tcccgccgtc ccgaccctg 27360

tcccgaccct cgtcccggcg gtcctgccct ggacactctc cggcaggagc accgcggcgc 27420

tgcgcgccca ggccgccaga cttctcacca cccagggcca ggacggtgcg accgaacccg 27480

ggcgtcccct cgacatcggc tactcactgg ccaccacccg cgcagccctt gagcaccgcg 27540

cggtgctcct cgggcgtacg gaggacgact ttgccgccgc cctctcggcg ctcgccgagg 27600

gtgcggagtc cgcaggcctg gtacagggca gggtgaccga gggcgggctg gcgttcctgt 27660

tcacggggca ggggagtcag cggctgggga tgggccgtga gctgtatgag gcgtatccgg 27720

tgttcgcgga tgcgctggat gcggtgtgtg cccgtcttga actgcctttg aaggatgttc 27780

tgttcggggc ggatgcgggt ctgctggacg agaccgcgta cacgcagccg gcgttgttcg 27840

ccgttgaggt ggcgctgttc cggttggtgg agagctgggg tgtgaagccg gacttcgtgg 27900

ccgggcattc gatcggtgag atcgcggccg cccatgtggc gggggtgttc tcgctggagg 27960

atgcgtgcgc gctggtgtcg gctcgtgggc ggttgatggg cgcgctgcct gcgggtggcg 28020

tgatgatcgc ggtccaggcg tcggaggccg aggtcctgcc gctgctgacc gaccgggtga 28080

gcattgccgc gatcaatggt ccccagtcgg tcgtgatcgc gggtgacgag gccgacgcgg 28140

tggcgatcgc agggtccttc gccgaccgca agtccaagcg gcttacggtc agtcacgcct 28200

tccactcgcc gcacatggac ggcatgttgg aggacttccg gctcgtggcg gagggcctgt 28260

cgtacgaggc cccgcgcatc ccggtcgtct cgaatctcac cggtgctctc gtctccgatg 28320

agatgggctc ggctgagttc tgggtgcggc acgtccgcga ggccgtccgt ttccttgacg 28380

gcatccggac gctggaagcc gctggcgtga ccaagtacgt cgaactcggc cccgacggcg 28440

tgctgtcggc gatggcccag gactgcgtga gtggcgaggg ctccgtcttc atccccgtgc 28500

tccgcaaggc acgccccgag gccgagagcg tcaccaccgc cctcgccacg gcccacgtcc 28560

acggcatccc cgtcgactgg caggcgttct acgccggaac cggcgcccag cgcgtcgacc 28620

tccccaccta cgccttccag cacgagcgtt actggctgga gcccgccacc ggcggagccg 28680

gtgatgtgag cggagccggg ctcgacccgg ccgggcatcc cctgctcggc gcggccgtca 28740

ccctggccgg ctcggacagt gtgctgttca ccggtcggct ctcgctccgc acgcagccct 28800

ggctcgccga ccacaccgtg tccggtacca ccgtgctgcc gggcgccgca ttcgtcgaac 28860

tcgccgtgcg tgccggtgac caggcaggct gcgagcgggt cgaggcgttg gtgctcgatg 28920

cgccgctcgc cctgcccgcg gagggcgccg tacgcgtcca ggtgctcgtc gaggcgcccg 28980

acgagcaggg ccgccgtccc ttcaccgttt cctcccagcc ggagaccgcg ccggccgaca 29040

ccccctgggg gcggcacgcc cggggcgtgc tcgcgcccac ggcccccgca ccgtcgttcg 29100

atctggcgca gtggccgccc gccggggccg aggccgtgga catcacggac ctctacgcgt 29160

cccacgacac ccctggcgcg cacgggcccg agcgcggtgg cctgttccgt gccgtggagg 29220

ccgtctggcg ctgtgacggt gacctcttcg ccgaggtgcg tctgcccgag ggcggcccgg 29280

acgcacaggc cttcggcctg cacccggcgc tgctcgacgc cgccgcgcac gcggcctcgg 29340

tactggacga gcagcacgga acgggggcag ggctgggcac gtggtccgat gtgactctgc 29400

acgccgtggg cgccggcgcc ctgcgcgtac ggatacggtc ggccctcgac ggcactgtgg 29460

gcctggacct cgcggacgac ctgggtgaac cggtggcgac cgtgggcggg ttgactccgc 29520

gacccttcgc gcaagcgggt tcaggtggac aggttgtcca gcatgacgcg ctgttccagc 29580

tcgactgggt gcggctgccg ctcgccgacc gctcgtccgc tcccaccggg gagtgggccg 29640

tactcggctc tgccgacggg ttcgcggacc tggaggcgct gggcgcagcg gtcgacgcgg 29700

gtgctcccgt accgccgtac gtcgtcgtcc ccttggagcg gcaggccacc ggcaacgggt 29760

cggacgccct gcacgaggcc gtgcaccggg cgctcgccct ggtgcggtcc tggctggacg 29820

accagcgctt cgagacctcg cgcctcgtgg tcctgacccg aggcgcggtc gccgggcccg 29880

gcgaaggcgt cgaggacctg ccgcatgccg cggtgtgggg cctggtgcgt tcggcggaga 29940

cggagaaccc cggccgtttc gttctcgccg acgtagacgt agacctcgac gcggacttgg 30000

gctcaggcgt gggcctcgcc gccgtactcg cctccggtga gccggagttg ctgctgcggg 30060

acggagtcgt acacgccccc cggctgaacc gggcccgtac cgccacctcg tccgacgccc 30120

ccggcatcga tccggccgga accgtcctga tcaccggtgg gtccggcacg ctcgccggta 30180

tcgtcgcccg gcacctggcc accgcccacg gtgtgcggcg tctgctgctg ctgagccgca 30240

ggggcgccga tgcccccggt gccggtgaac tgaccgctga gctggccggg ttgggcgcgc 30300

aggtctcgtg ggcggcgtgt gacgcgggtg accgcgacgc gctcgcggcc gtactggccg 30360

ccgttcccgc agcgcacccg ctcaccgcgg tcgtccacac ggccggtgtc ctcgacgacg 30420

gcgtgatcgg ttcgctcacc ccggaacgtc tcgacacggt ccttcgcccg aaggccgatg 30480

ccgctctcca cctgcacgaa ctgacccgcg acctgcccct gaccgccttc gtcctcttct 30540

ccgcgatcgc cggaaccctc ggcagtgcgg gtcaggccaa ctacgcggcc gccaacgtct 30600

tcctggacgc tctggcccag caccgccatg accaggacct gccggccacc tcgctcgcct 30660

ggggcctgtg ggccgatgcc agcgggatga ccggcggcct cgacgaggcc cagctgcggc 30720

gcatggagca gcacggcatg ggcacgctct ccgccaccga cggcatggcg ctgttcgacg 30780

ccgccctcgc cgccggccgg ccggtcctcg tcccggcccg tctgcacctc cccggcctgc 30840

gcaatgccgc cgggccgggc ccggtggctc cggtgttccg gtcgctcctg ggtgcctcgg 30900

gccgccgggc cgcgcggacc cgtaccgacg gcggcacccc gctcgccgag cggctgaccc 30960

gcctcgccgg tcccgaacag gaccgggcgc tgctcgatct cgtacgggca caggtcgcat 31020

ccgtactcgg ccacgcctcg gccgaacagg tggaccccgc acgcgcgttc aaggatctgg 31080

gcttcgactc cctgaccgcc gtcgagctgc gcaaccggct gggcgccgcc accggactcc 31140

ggctgccgac cacgctcgtc ttcgatcatc cgacgcccac cgcgctcgtc cggcacttgc 31200

gtacggacct tctcggcgcc gcgccggacc ccggagccga cgccccgggc ctgcccgcgc 31260

gcgtcggcct cgccgacgac ccgatcgcca tcgtggccat gagctgccgc taccccggcg 31320

gtgtccgcac ccccgaggag ctgtggcggc tcgtcgagac cggtggcgac gcgatcgccg 31380

gactcccggg caaccggggg tgggacaccg acgcgttgca cgccgacgag gacggccgga 31440

ccttcgcggg cggcttcctg tacgacgccg actcgttcga cgcggacttc ttcggcatct 31500

cgccgcgcga ggcgctcgcc atggacccgc agcagcgact gctgctcgaa acctcctggg 31560

aggcgatcga gcgcgccggg atcgacccgt cgtcgctgcg cggcagccgg gccggtgtct 31620

tcgtcggcgc cgcctacagc ggctacgacg cgcaattgga gcagtccgga gtggacggtg 31680

tcctcggcca tgtgatgacc ggcaatgcgg gcagtgtcat gtccggccgt gtgtcctacg 31740

cgctgggcct ggagggtccg gcggtcacgg tcgacacggc gtgctcgtcc tcgctggtcg 31800

ccctgcactg ggcgatccag gccctgcgca acggcgaatg ctcgctggcg ctcgccggtg 31860

gtgtgacggt gatgtcgacc ccgggcacct tcagcgagtt cagccagcag ggcggcctgt 31920

caccggacgg ccggtgcaag gcgttcgcgt cggccgcgga cggtacgggc tggggtgagg 31980

gtgtcgggat gctgctggtg gagcggctgt ccgatgcccg taggaatggg catccggttc 32040

tggcggtggt gcgtggttcg gctgtcaatc aggacggtgc gagcaatggt ctgacggctc 32100

cgaatggtcc ttcgcagcag cgggtgatcc gtgcggcgtt ggcgagtgcg ggtctgtcgg 32160

ccgctgatgt ggatgtggtg gaggcgcacg gtacggggac gaagctgggt gacccgatcg 32220

aggcgcaggc gctgctggcg acgtacgggc aggaccggcc cgatggccgt ccgctgtggt 32280

tgggttccat caagtccaac atcggtcaca cgcaggccgc cgccggtgtc gcgggcatca 32340

tcaagatggt catggcgatg cggcacgggg tgctgccccg gaccctgcac gtcgacgagc 32400

cgacctcgca tgtggactgg tcggcgggcg aggtgtccct gctgtcggag tcggccgaat 32460

ggccgctcac cgagcggccc cggcgagccg gagtgtcgtc cttcggcatc agcggcacca 32520

acgcccacac catcatcgag caggcgccgg agaccgggac cgaggcggag ccgtcggcgg 32580

agaccctcac gcacgggacc gtgccctacg tcctctccgc caagagctcc gacgctctcc 32640

gcgcccaagc gcggcagctg cttgccgtgg tggaagccgc cgagagcccc cgagtcgccg 32700

atctggccta ctcgttggcc accagtcggg ccggtctcga tcaccgcgcg gcgctcgtcg 32760

ccgacgaccg ggagaacctg acgcgggcgc tcgcggccct ggcggcggac gagcaggtgc 32820

ccggcctggt gcggggcacg gccaccggtg gcggcctcgc cttcctgttc acggggcagg 32880

ggagtcagcg gctggggatg ggccgggagc tgtacgagac gtatcccgtc ttcgcgcggg 32940

ctctcgacgc ggtggacgca cgcctggaac tgcccatgaa ggaggtgctg ttcggcgcgg 33000

acgcggatct gctgaacgag accgcccaca cgcagccggc tctcttcgcc gtcgaggtgg 33060

cgctgttccg tctgctggag tcgtggggcg tgcggcccga cgtcctggcc gggcactcga 33120

tcggtgagat cgccgcggcc catgtggccg gggtgttctc cctggacgat gcgtgcacgc 33180

tggtcgaggc tcgcggtcgg ctcatgcagg cgctgccgac cggcggcgtg atgatcgccg 33240

tccaggcgtc ggaggacgaa gtcctgccgc tgctgaccgg ccaggtgagc attgccgcga 33300

tcaacggccc ccagtcggtc gtcatcgcgg gcgacgaggc cgacgcggtc gcgatcgccg 33360

agtccttcac cgaccgcaag tccaagcggc tcaccgtcag ccacgccttc cactcgcccc 33420

acatggacgg catgctcgcc gacttccgca aggtcgccga gggcctcgtc tacgagaacc 33480

cgcgcatccc catcgtctcg aacctcaccg gcactctcgt caccgacgag atggcttcgg 33540

ccgacttctg ggtccgccac gtccgcgagg ccgtccgttt cctcgacggc atccgcgcgc 33600

tggagagccg cggggtcacc acctacatcg aactcggccc cgacggggtc ctctccgccc 33660

tcgcccagga ctgcctcacc gccgggaccg ggaccgggac cgcgatcttc gctcccgtac 33720

tccgggcggc ccgtcccgag gccgagagcg tcaccaccgc cctcgccacg gcacacgtcc 33780

acggcacccc cgtcgactgg cgggcgtact tcgccgggac cggtgcccgg cgcgccgacc 33840

tccccaccta ccccttccag ggcaggcgct actggcccga agccgccgcc ccgagcggtg 33900

cggcggccgg actcggggac caggcggtcg acgcgcgctt ctgggacgcg gtcgagcggg 33960

cggacctggg ctccctgatc ggtgggccgg agatcgacgg ggaccagccg ctcagctccg 34020

tactgcccgc cctctccgac tggcggcgca accagcaggc gcagtcgcag gcggacgccc 34080

ggctctaccg catcgcgtgg cagccgtggt ccggggccgg ccggggcaca cccgcgggta 34140

cctggctggt ggccgtgccg gcgccgtacg cggacgatcc gtgggtccgt gcgctgaccg 34200

accgcatggc cgagggtggc gcggaggtcg taccgctcac gctcgatgtc gccgacagcg 34260

acccggcgtc gctgcgcgcc cggctggacg agcggctgcg cgaggcggtg ggcgacggcc 34320

cggtggccgg tgtcctgtcc ctgctcgcgc tggacgagcg gccccacccc gaccacccga 34380

gcgtgcccgt aggactggcc ctcaccagcg ccctcacctc cgtgctcacc ccggtgctca 34440

cggaaccgga cccggaaggc ggggcgagcg gaggcatcga agcaccgctg tggtgtgtca 34500

cgcgtgacgc cgtcgcggca gccggtggtg acgaactcgg cggcgccgcc caggcgcagg 34560

tctggggcct cggccgcgtc gtcgccctgg agcaccccga ccgctggggc ggtctcgtcg 34620

acctcccggc ggtatgcgac gaccgggtcc tgtcccggct gatggcggtg ctcgcaggat 34680

ccggtgacga ggaccaggtg gcggtccgta cctccggcac cctcgtacga cggctcctgc 34740

gggccgcccc gacgagcgtg ccgtccgcac cctggacccc gcgcggcacg gtgctcgtca 34800

ccggcggcac gggcgccctc ggccgccatg tggcgcgcca cctcgccgag cggggcgccg 34860

aacggctcgt gctcgtcagc cgccggggcg ccgacgcgcc cggtgcggcc gagaccgagg 34920

cggaactctc cgcgttcggc gcggccgtga ccctcgtggc ctgcgacgtc gccgaccgcg 34980

atgcgctcgg aacgctcgtc gcgcggctcg ccgccgacgg cactccggtc cgtgccgtgg 35040

tgcacgccgc cggtgtctcg cagccgccag gtacgggaac ggacctcccc gggttcgccc 35100

gtgtcgtggc cgcgaagacg gcgggagccg tccacctcga cgcgctgttc gacgcgccgg 35160

actccctcga cgcgttcgtc ctcttctcct ccatcgccgg tgtctggggc agtggcggcc 35220

aaggggccta ctccgccgcc aacaccttcc tcgacacgct cgccgaacgg cgccgggccc 35280

gcggtctcgc cgccacggcg atcgcctggg gaccgtgggc cgacggcggc atggccaccg 35340

agggcgacgc ggaggagcag ctgagccgac gcggcctgcc gcccatggac cgggcgacga 35400

acctgctggc gctggagcgt gccgtcgcgg gccgggaggc ggcgctgacc gtcgccgacg 35460

tcgactgggc gcgcttcgca cccgtgttcg ccgcggcccg cccccgcccg ctcatcggcg 35520

acctgcccga ggtacgggac gcactgcgcg gggacacccc ggccggggaa ggaccggccg 35580

agaccgcttc ctccgccgta ctccggaggc tgacggaact caccggggcg gaccgggaaa 35640

cggccctcct cgacctcgtg cgcgagcacg cggcaacggc cctgggccac acgtccgccg 35700

acgcggtcgc ggccgaacgg gccttcaagg acctcggctt cgactcgctc accgcagtcg 35760

aactgcgcaa ccgcctcggc gccgcgtgcg gcctgcggct gccctccagc ctcgtcttcg 35820

actaccccaa cccgcaggcg ctcacccggc acctgctgca caccctcttc cccgaagggg 35880

cgggcgggcc ggacgtaccg gctctggaca ccgaccccca ggaagcggaa ctgcgccgga 35940

cgctcgccgc catcccgctg ggccggatcc gcgaggcagg gctcctggac acgctgctcc 36000

ggctcgccgg acccgacacc cccgctcccg ccacgagtac cgccgacgag agcgagtcca 36060

tcgacacgat ggatctccag gacctcctcg acctggcgct cgacggcggc ggcgatcccg 36120

acggcctcaa cggcctcgac agcctcgacg gccccagtgg caacgacaac gacagcaacc 36180

gattctgacg tgcccgaagt gcggagtaag tgatgacaac ccccaacgaa aaagtcgttg 36240

aagcgctgcg ggcctccctc aaggaaaccg agcggctgcg ccgccggaac caggagctca 36300

ccgacgccgc gcgcgagccc atcgcgatcg tcggcatgag ctgccgcttc ccgggcggag 36360

tcagctcgcc cgaggacctg tggagactcg tcgagagcgg tggcgacgcc atctcgggct 36420

tccccgtcaa ccgcggctgg gacatcgagt cgctgtacga ccccgatccg gaccacgagg 36480

gcaccaccta cgccgcgac ggcggcttcc tccacgaggc ggccgacttc gaccccgcgt 36540

tcttcgggat ctccccgcgc gaggccctcg ccatggaccc gcagcagcgg ctgctcctgg 36600

agaccacctg ggaggtcttc gaacgagccg gaatcgatcc cgcgtcgctg cgcggcagcc 36660

gggccggcgt cttcgtcggc gcgtccgcca acgcctacgg agccggctcc cacgaccttc 36720

ccgacggcgt ggagggacac ctcctcaccg gcaccgcgtc cagtgtcctg tccggccggc 36780

tcgcctacgt cttcggcctg gagggccccg ccgccaccat cgacacggcg tgctcgtcct 36840

cctccgtcgc cctgcacatg gccgtccagg cgctgcgcca gggcgagtgc tcgctcgcgc 36900

tggccgcggg cgtcaccgtc ctcgcgggcc cggacgtctt cgtcgagttc agccgccagc 36960

gcggcctgtc gcccgacggc cgctgccggt ccttcgccga gtcggccgac ggcaccggct 37020

ggtcggaggg cgccggcgtc ctcctggtgg agcgcctctc cgacgcccgc cgcaacggcc 37080

accacatcct cgccgtggtc cgcggctcgg ccgtcaacca ggacggcgcc agcaacggcc 37140

tgaccgcccc caacgggccc gcccagcaga aggtcatccg ccaggccctg gagagcgccc 37200

ggctgacccc cgcggacatc gacgcggtcg aggcccacgg caccggcacg accctcggcg 37260

accccatcga ggcgcaggcg ctcctcgcca cctacgggca agggcgcacg gacggccggc 37320

cgctgtggct cggctccttg aagtcgaacc tcggccacac ccagaacgcc gccggtgtcg 37380

ccggcatcat caagatggtc atggcgatgc ggcacggggt gctgccccgg accctgcacg 37440

tcgacgagcc cacctcgcac gtcgactggt cgacgggcgc ggtggcgctg ctgaccgagc 37500

cggtggagtg gccggagacc gggcgcccgc gccgggtcgg cgtctccgcc ttcggcgtca 37560

gcggcacgaa tgtgcacacg atcatcgagc aggccccggc ccctgccccg gcccccgtcg 37620

cggacgacac atcggaaccg gcgcccgccg cccggccgaa ggcgctgccc tggctcctct 37680

ccgcgaaggg ccgggacgcc ctgcgcgacc gggccgcaca gctgctcgcg tacgccgagg 37740

aacaccccga cctgcggccg gtcgacatcg ccgggtcgct ggcggtgggc aggccgtcct 37800

tcgaggaccg cgccgcggtg gtcgccgccg accgcgaggg gctgctggcc ggcctcgcgg 37860

cactggcgga cggcggctcg gcgacgggtc tcgtcaaggg gtcgtcgcag ctcgtgggga 37920

agctggcgtt cctgttcacc gggcagggga gccagcggct ggggatgggc cgtgagctgt 37980

acgagacgta tcccgtcttc gcgcaggcct tggacgcggt gtgtgagcgg ctggaactac 38040

ccctgaagaa cgtgctgttc gggacggaca gcgctgcgct ggacgagacc tcgtacacgc 38100

agcctgctct cttcgccgtt gaggtggcgt tgttccggct cgtggagagc tggggcctga 38160

agccggactt cctggccggg cattcgatcg gtgagatcgc ggccgcgcat gtggccgggg 38220

tgttctcgct ggacgacgcg tgcgcgctgg tgtcggctcg cggccggttg atgggggcgc 38280

tgccgggcgg tggcgtgatg atcgcggtcc aggcgtcgga ggacgaggtc ctgccgctgc 38340

tgaccgatcg cgtgagcatt gccgcgatca acggtccgca gtcggtcgtg atcgcgggtg 38400

acgaagccga tgcggtagcc atcgccgagt ccttcgcgga ccgcaagtcc aagcggctca 38460

cggtcagtca cgcgttccat tcgccgcaca tggacggcat gttggaggac ttccgggtcg 38520

tggcggaggg tctgtcgtac gaggctccgc gcatcccggt cgtctcgaac ctcaccggcg 38580

ctctcgtctc cgacgagatg ggctcggccg acttctgggt ccgccacgtc cgcgagaccg 38640

tccgcttcct ggacggtatc cgcaccctgg aagccgctgg cgtcaccaag tacgtcgaac 38700

tcggcccgga cggcgtgctg tccgccctgg cccaggactg cgtgagcggc gaggactccg 38760

tcttcatccc tgtactccgc aaggcacgcc ccgaggccga gacggtcgcc accgccctcg 38820

cctcggccca cgtccacggc atccccgtcg actggcgggc gtacttcgcc gggaccggcg 38880

cccagcgcgt agacctcccc acctacccct tccagcgcca gcgctactgg atcgagccgg 38940

gcggccgtgc cggagacgtg ggcgcggccg ggctggagga ggcggggcat ccgctgctgg 39000

gtgcggccgt accgctcgcc gactccgagg gcttcctctt caccgggcgg ctcggtcgca 39060

cctcgcaccc ctggctggcc gatcacgcgg tcatggacac cgttctgctc cccggcacgg 39120

ccttcgtcga cctcgcggtg cgcgccggtg accaggtcgg atgcgatgtc gtcgaggagc 39180

tgacgctgga agcgccgctg gtgctgcccg agcgcggtgc cgtccagata cagatgcacg 39240

tcggcgcgcc cgacgcggac ggtacgggac ggcggacgtt caccctgtcc tcgcgtacgc 39300

aggacggcgc ggccgacgaa ccgtggacgc ggcacgccgg cggcgtcctc gcgcacggcg 39360

cggcgcaacc ggccttcgcg ccggtccagt ggcccccggc gggtgccgag ccgatcccga 39420

cggagagcct gtacgcggac ctggccgagg tcggcatggg atacggaccc gcgttccgcg 39480

gcctcacggc cgcctggcgg cacggcgaga gcgtctacgt cgaggtcgcg ctccccgagg 39540

aaaccgcctc cacggcacgg gacttcggcc tgcaccccgc cctcctggac gcggcgctgc 39600

acgcgctggg tctcggcgta ctgggtggcg tcgagggtga agggcggctc cccttcgcgt 39660

ggagcggtgt gaccctgcac gcggccggag cggacgcgct gcgcgtgcac ctcgctccgg 39720

cgggcgccca cggcgtacgc ctggagatcg cggacgccgc gggcgcacct gtcgcgaccg 39780

tcgactcgct cgtcctgcgg accgtatcgg aggagcaggt acgcgccgcg cgcaccgcgt 39840

accacgagtc ggtgttccgg gcggagtgga cggccctgcc gaccgccgcc gaatccgcgg 39900

ccacgcatgg ccgttgggcc gtgctgggag cggcggacgc gggcgattcg ccgcgcgacg 39960

cgctggtgaa cgggctgctc ggccacctgc ccggcgaggt cgcgcgctac gccgacctgg 40020

ccgagctggc ggcggccgtc gaggccggag cggccacgcc ggacgccgtg ttcgccgcgt 40080

acgcgcggtc cgatgacgac ggaccggccg caccggacgt gtccgcaccg gacgtgtccg 40140

cgcaggcggt gcacgcggcc acccacgacg ccctcgcact cgtccagacg tggttcggtg 40200

aggagccctt cgccggggac cggttcgccg ccacccgcct ggtcgtgctc acccggggcg 40260

cggtcgcggc gggcgacggc gacacggtca ccgaccccgc acacgcggcc gtctggggtc 40320

tgctgcgctc cgcgcagtcc gagtaccccg accggctgct gctgatcgac accgacgggg 40380

tcgaggactc cgtacacgcc ctgcccgccg tgctcgccgt cggagagccg caactcgccc 40440

tgcgtgcagg ctccgtacac gcgctccggc tcgcccgcgt ggccgccgcg acgccggagg 40500

acgccgccgc tccgacgcag tacgcgcccg gatcgacggt gctgatcacc ggcgcgggcg 40560

gcatgctcgg cggtctgatc gcccgccgtc tcgtcgccga acacggcgta cggcacctgc 40620

tgctggtggg ccgccgcggc gccgccgctc ccggagcgga acagctgagc gccgaactgg 40680

ccgaggcggg cgcctcggtg acctgggccg cgtgcgacgt cgccgaccgg gacgcccctct 40740

cggccgtact gcacgcgata cccgccgagc acccgctcgg cgcggtcgtc cacaccgctg 40800

gtgtgctgga cgacggtgtg atcgcctcac tgaccccgga gcggctctcg gccgtgctgc 40860

gccccaaggt cgacgccgcc tgcaacctcc acgagctgac ccggcacctc gacctcacgg 40920

cgttcgtgct cttctcctcc atcggcggcg tcttcggcgg cccgggacag ggcaactacg 40980

cggcggcgaa cgtgttcctc gacgcactcg cccagcaccg ccgctcccag ggactcgccg 41040

ccacctccct ggcctgggcc ctgtgggccg acagcacggg catggccggc agcctcgacg 41100

aggccgacat cagccggatg cggcggggcg gcctgccccc gctgaccacg gccgagggcc 41160

tggaactgtt cgacctcgcc caccgcatcg acgaggccgc accggtcctg atgcgcgccg 41220

acctgaccgc cctgcgcacg caggcccagg ccggcacgat gtcgccgctg ctgcgcggtc 41280

tcgtacgggt ccccgcgcgc cgcagcgcca gtggcgcggc cggtacgggc ggtgagtccg 41340

gactgcgcga gcgcctcgcc ggactctcgg ccgccgaacg ggaccgtacg ctgctcgacc 41400

tcgtccgcaa gcaggtcgcc gcggccctcg gctaccccgg accctccgcc gtcgagcccg 41460

gccgctcctt caaggaactc ggcttcgact cgctcaccgc cgtcgaactg cgcaacctgc 41520

tcggcgacgc caccggccgc cgcctccccg ccaccctcgt cttcgactac ccgacggcga 41580

ccgccctcgc cgggtacctc cgcgaggaga tcatcggaga cctggcggac gccgtcaccg 41640

ccccggccct cgtgccgtcc gcggccgtgg cgggcgcggg cgcgggcgcg gacgacgacg 41700

atccgatcgc gatcgtcgcc atgagctgcc ggttccccgg agggatcgca tccccgagg 41760

acctgtggca gctgctcgtc accggccgcg acggcatcac gggcttcccg gcggaccgtg 41820

gctgggacct cgacagcctc tacagcgacg accccgaccg cgagggcacg agctacgccc 41880

gcgagggcgg attcctgcac gaggccgccg agttcgacgc ctccttcttc gggatctcgc 41940

cgcgcgaggc cctcgccatg gacccgcagc agcggctgct cctggagacc acctgggaga 42000

cgttcgagcg cgcgggcatc gacccgacca gcctgcgcgg cagccggacc ggcgtgttcg 42060

tcggctccaa cgcccaggac tacctccagc tctggctgaa cgacgcggac ggcctcgaag 42120

gacacctggg caccggcaac gcggccagcg tcgtctccgg ccgcctctcc tacaccttcg 42180

gcctggaggg cccggccgtc acggtcgaca cggcctgctc gtcctccctc gtcaccctgc 42240

acctggccgc ccaggccctg cgccgcggcg agtgctccat ggcgctcgcc ggcgcggtca 42300

ccatcatgtc cacgcccggc gcgttcaccg agttcagccg ccagcgcgga ctcgccgccg 42360

acggccgcat caaggcgttc gccgccgccg ccgacggcac gagctggtcc gaaggcgtcg 42420

gcctgctgct cgtcgagcgg ctctcggacg cacggcgcaa cggtcacccg gttctggcgg 42480

tggtgcgggg caccgccgtc aaccaggacg gcgcgagcaa cggcctgacc gcgccgaacg 42540

gcccgtccca gcagcgcgtc atccgcgagg cgctggccga cgcgggcctg tcggccgccg 42600

aggtggatgc ggtcgaggcc cacggcaccg gcacgaccct cggcgacccc atcgaggcgc 42660

aggcgctcct cgccacgtac ggccagggcc gcccggacga ccagccgctg tggctcggct 42720

ccgtgaagtc caacatcggc cacacccagg ccgtggccgg agccgccggc atcatcaaga 42780

tggtcatggc gatgcgccac ggcgtactgc cgcagaccct gcacatcgac gagccgacgc 42840

cgtacgtgga ctggtcggcg ggcgacatcg ccctgctgac cgagcagcgg gcgtggccgg 42900

agaccggccg cccgcgcagg gcgggcgtct cctcgttcgg ctacagcgga accaacgcgc 42960

acgccgtcat cgagcaggca ccgcagaacg cgatggagcg gaccccgcag ggcgacaacc 43020

tgccggcccg caccccgcg acgcggaccc tcccggtgct gccgctgctc gtctccggcc 43080

gcacggcgcc ggccctgcga gcccaggcgg aacgcctgcg accggccgcg accgccctcg 43140

cgacgggcac ggtaacgaac tccggagctt tggaagcact cgacctgggc tactccctgg 43200

ccacgagccg cgccgcactg gaacaccggg cggtcctgat cggcaccccg tcggacggcc 43260

aggcactggc ctcgcgactc gacgccctgg cggcgggcga gcaggtgccc ggcctggtgc 43320

agggcacggc ttccggtggc gggctcgcct tcctgttcac gggacagggg agccagcggc 43380

tggggatggg gcgcgagctg tacgagacgt acccggtgtt cgcggaggcg ttggatgcgg 43440

tgtgcgcccg gctcgaactg cctttgaagg aggtgctgtt cggggcggat ggcgctgcgc 43500

tggatcagac ggcggtgaca cagccggccc tcttcgccat tgaggtggcg ttgttccggc 43560

tggtcgagtc gtggggtctg aggccggact ttgtggcggg tcattcgatt ggtgagatcg 43620

ccgctgcgca tgtggcgggg gtgttctcgc tggaggacgc ctgcaggttg gtcgaggcgc 43680

gtgggcgtct tatgcaggcg ctgcctggtg gtggcgtgat gatcgcggtc caggcgtcgg 43740

aggatgaagt cctgccgttg ctgaccgatc gcgtgagcat tgccgcgatc aatggtccgc 43800

agtcggtggt gatcgcgggt gacgaggccg acgcggtggc catcgcggag tccttcacgg 43860

gccgcaagtc gaagcatctg gcggtcagcc acgcgttcca ttcgccgcac atggacggca 43920

tgttggagga cttccgggcc gtggcggagg gcctgtcgta cgaggctccg cgtattgcgg 43980

tggtgtcgaa tctgacgggt gcgttggtct ccgacgagat gtcgtcggct gagttctggg 44040

tgcgtcatgt ccgtgaggcg gttcgcttcc tggacggtat tcgggctttg gaggctgctg 44100

gggttacgac gtatgtcgag cttggccctg ggggtgtgct gtcggcgctg gcgcaggagt 44160

gtgtcagtgg ggacggtgct gctttcgtgc cggtgctgcg ttctggacgt tccgaggccg 44220

agaccgtggt gaccgcgctg gctcaggcgc atgtgcgggg tgtggaggtc gactgggcgg 44280

cgttcttcgc cgggaccggt gctgagcgga tcgatctgcc gacgtacgcc ttccagcgcc 44340

agcgctactg gccggagacc gtgctgtcga ccgtgggccc ggtcgttgcc gaggccgtcg 44400

atgcggtgga cgcccggttc tgggatgcgg tggagcggga ggatctcgcg tcgcttgtcg 44460

cagagctgga cgtggacgag acgcctctcg gcgaggtcgt tcccgcgctg tcggcgtggc 44520

gtcgggagcg gcgtgcccag tcggaggtgg acggttggcg ctaccgggtg tcgtggaagc 44580

cgctggctga tgcttcgacg gcgcggttgt ccggctcttg ggtggtggtg tcgcccgata 44640

agggtgtgga tgactcggct gtggtcgccg gtctggctgg gcgtggtgct gaggtccgtc 44700

gggttgtggt cgaggcgggt gtggaccgtt cggcgctggc tgggttgctg gccgatgcgg 44760

gttctgctgc gggtgtggtg tcgcttctcg ggctggatga gtctgagggg ctgctgggga 44820

ctgttggttt ggtgcaggcg ttgggtgatg ccggggtgga ggcgccgttg tggtgcctga 44880

cccgtggtgc tgtctccgtc ggtcgttcgg atcggcttgt gtcgccggtg caggcgcagg 44940

tgtggggtct gggccgggtt gccgccctgg aggttccgga gcattggggc gggctggttg 45000

acctgccgga agtgctggat gagcgggctg tggcccgctt ggtcggtgtg cttgcgggtt 45060

ccggcgaaga tcaggtcgcg gttcgttcgt ctggtgtgtt cggtcgtcgt ttggtgcgtg 45120

caccgcgggc cgagggtgct gcggcgtgga caccgaccgg cactgttctt gtcaccggtg 45180

gtacgggtgt gctgggtggc cgggtggcgc gttggctggc gggggcgggc gctgagcgtc 45240

tggtgctgac cagtcgtcgt ggtccggatg ctccgggtgc ggctgagctg gtggaagagc 45300

tgaccaccgg cttcggggtg gaggtttcga tcgtcgcgtg tgacgcggct gaccgtgacg 45360

ccctgcgcgc cctgctctcc gctgaggccg ggactctgac cgctgtgatc cacacggccg 45420

gtgtcctgga cgacggcgtc ctcgacgcac tcaccccgga ccgcatcgac agcgttctgc 45480

gcgccaaggc cgtctcggca ctcaacctgc acgaactgac ggccgagctt gatatcgagc 45540

tgtccgcctt cgtcctcttc tcgtcgatga gtggcacggt gggtgcggcc ggtcaggcca 45600

actacgcggc cgccaacgcc ttcctggatg ccctggccga gcagcggcgc gccgatggtc 45660

tcgcggcgac ctcgctcgct tggggtccgt gggcggaagg cggcatggcc gccgatgcgg 45720

cgctcgaagc ccgtatgcgc cgcggcggag taccgcccat ggacgcggag cttgcccttt 45780

cggctcttcg gcaggccatc ggttccgccg atgccgctct gaccatcgtg gacttcgact 45840

gggcacggtt cgcgcccggc ttcaccgccg tgcgagccgg caacctgctc gccgaactgc 45900

ccgaggcggc ggccgtcatg cgcggcccgg agaacgcgga cagccgcccg gaacacgccg 45960

actcgtcgct cgccctgagg cttcagggca tggcccaggc cgaccaggag cctttccttc 46020

tggagctcgt gcgtgcacag gtcgccgagg tgctgggaca ctccggcgcc gaggacatcg 46080

aggcgggacg cgcgttcagg gagatcggct tcgactcgct gaccgccgtc gagctgcgca 46140

accgcctcgg ggcggctgcc gagctgcggc tcccggccac gctcgtctac gactacccga 46200

caccggcggc cctcgccgtc cacctccgta ccgaactgct cggcaagcag gtcgtcgtgt 46260

ccggtccggt ctccaaggtc gttgacgacg atccgatcgc gatcgtctcg atgagctgcc 46320

gcttccccgg tggcgtgcgg accccggaag acctgtggga actgctgtcc accggcggcg 46380

acgccatctc ggatcttccc ctggaccgtg gctgggacat cgacgcgctg tacgacgccg 46440

atcccagcac acagggcact tcgtacgccc gcgcgggtgg cttcctctac gacgccgccg 46500

acttcgacgc ggacttcttc gggatctcgc cgcgcgaggc cctcgccatg gacccccagc 46560

agcgactgct cctggagacg tcctgggaag ccttcgagcg ggcgggcatc gaccccgaga 46620

cgctccgggg cagccaggcc ggtgtcttcg tcggcaccaa cggccaggac tacctctccg 46680

tactgctgga ggagcccgaa ggcctcgaag gccacttggg caccggcaac gcggcgagcg 46740

tcgtctccgg tcggctctcg tacgtgttcg gcctggaggg tccggcggtc acggtcgaca 46800

cggcgtgctc gtcctcgttg gtcgccctgc actgggcgat ccaggccctg cgcaacggcg 46860

aatgctcgct ggcgctcgcc ggtggtgtga cggtgatgtc gaccccgggc accttcatcg 46920

agttcagccg tcagcgtggg ctcgcggagg acggccgtat caaggcgttc gcggcggccg 46980

cggacggtac gggctggggc gagggcgtcg gcatgctcct ggtggagcgg ctgtccgacg 47040

ccgagcggaa cgggcacccg gtcctggcga tcgtgcgggg ctcggcgatc aaccaggacg 47100

gtgcgagcaa cggcctcacc gcccccaatg gcccctcgca gcagcgcgtg atccgtgcgg 47160

cgctggcgag cgcgggtctg tccgccgccg acgtggacgc ggtcgaggcg cacggcaccg 47220

gtacgacgct gggcgacccg atcgaggcgc aggccctgct cgccacgtac gggcaggacc 47280

gcccggccga ccggcctctg cagctcggtt ccatcaagtc caacatcggg cacacgcagg 47340

ccgcggccgg tgtcgccgga gtgatcaaga tggtgctggc catggagcac ggcgtgctcc 47400

cgcagagcct ccacatcgac gcaccgtcac cgcaggtcga ctgggaagcc ggtgacatcg 47460

cgctgctcac cgagcagcgg cagtggccgg agaccggacg tccccgccgg gcaggtgtgt 47520

cgtcgttcgg cttcagtggc accaacgctc acaccatcat cgagcaggca ccggcgtcga 47580

cggagaccga ccgggccgaa tccggctcgg tggaaccgga cttcgttccc ctgatgctct 47640

cggcgaagag cgacgtcgca ctccgggccc aggccgcaag cctgcgcgca cggctgatcg 47700

ccgcccccga catgcgcctg tccgacgtcg gctccacgct gacgaccggc cgctcggcgt 47760

tcgagccgcg ggcggcgctg gtggcagggg gccgcgaggg gctgctcgcg gggcttgagg 47820

cactggcgga cggcggttcg gcggcagggc tggtggaagg ttcgccggtg agtggaaagc 47880

tggcgttcct gttcacgggg caggggagtc agcgtctggg catgggcgt gagctgtacg 47940

aggcgtatcc ggtgttcgcg gatgcgctgg atgcggtgtg tgtccgtctt gaactgccct 48000

tgatggatgt gctgttcggg gcggatgcgg gtctgctgaa cgagaccgcg tacacccagc 48060

cggcgctctt cgccgttgag gtggcgttgt tccggctggt ggagagctgg ggtctgaggc 48120

cggacttcct ggcgggtcat tcgatcggtg agatcgcggc cgcgcatgtg gccggggtgc 48180

tgtccctgga cgatgcctgt gctctggtgg aggctcgggg gcggttgatg ggtgcgctgc 48240

ctgcgggtgg cgtgatgatc gcggtgcagg cgtcggagga cgaggtcctg ccgctgctga 48300

cggaccgcgt gagcattgcc gcgatcaatg gtcctcagtc ggtggtgatc gcgggcgacg 48360

aagccgacgc ggtcgcgatc gtggagtcgt tcacggggcg taagtcgaag cggctatcgg 48420

tgagtcacgc gttccattcg ccgcacatgg acggcatgtt ggaggacttc cgggtcgtgg 48480

cggagggcct gtcgtacgac gccccgcgca tccccgtcgt ctcgaacctc accggcgctc 48540

tggtcaccga cgagatgggt tcggcggact tctgggtccg gcacgtccgc gaggccgttc 48600

gcttcctgga cggcatccgg gccctggagg ccgcgggcgt gacgacgtac gtcgaactcg 48660

gccccgacgg tgttctgtcg gcgatggccc agggagtgtgt gaccgaaggt ggagcggcgt 48720

tcgttcccgt cctgcggaag gggcggcccg aggccgagac ggtgatggcc acccttggcc 48780

aggcacacgt cagggggcgtc gcggtcgact ggcattcggt ctacgggacc ggtgcccagc 48840

gggtcgatct gccgacctac tccttccagc gacagcggta ctggccggcg gcgtcttcga 48900

cggcaggtgg ttcggtcgac aggagcgtcg atgcggtgga cgcccggttc tgggatgcgg 48960

tggagcggga ggatctcgcg tcgctggccg cggagctgga cctggacgac gacgctccct 49020

tcagtgaact ggcccccgcg ctgtcggcgt ggcggcggga gcggcgtgcc ctgtcggagg 49080

tggatggctg gcgctatcgg gtgtcgtgga agccgctggc ggatgtctcg gcgtcggggt 49140

tgtccggctc ttgggtggtg atctcgcctg ctgggggtgt ggacgactcg gctgtggtgg 49200

gtgcgctggt tgggcgtggt gctgaggtcc gtcgggttgt ggtcgaggcg ggtgtggatc 49260

gttcggcgct ggctgggttg ctggccgatg cgggttctgc tgcgggtgtg gtgtcgcttc 49320

tcgggctgga tgagtctgag gggctgctgg ggactgttgg tttggtgcag gcgttgggtg 49380

atgccggggt ggaggcgccg ttgtggtgcc tgacccgtgg tgctgtctcc gtcggtcgtt 49440

cggatcggct tgtgtcgccg gttcaggcgc aggtgtgggg tttggggcgg gttgccgccc 49500

tggaggtccc cgagcgctgg ggcgggctca tcgatctgcc tgaggtgctg gatgagcggg 49560

ctgtgtcccg tctggtcggt gtgctttcgg gtggtggttc tggtgaggat caggttgcgg 49620

ttcgttcgtc gggtgtgttc ggtcgtcgtc tggtgcgtgc accgcgggct gaggggggctt 49680

cggcgtggtc tccgaccggc acggttcttg tcaccggtgg tacgggtgtg ctgggtggcc 49740

gggtggcgcg ttggctggcc ggggcgggtg ctgagcgtct ggtgctgacc agtcgtcgtg 49800

gtccggatgc tccgggtgcg gctgagctgg tcgaggaact ggccgggtcg ggggtcgagg 49860

tttcggtcgt cgcgtgtgat gcggccgacc gtgacgctct gcgcgccctg ctctccgccg 49920

aggccgggac tctgaccgct gtgatccaca cggccggagt tctggacgac ggcgtcctcg 49980

acgcgctcac cccggaccgc atcgacagcg ttctgcgcgc caaggcagtc tcggccatca 50040

acctgcacga actgacggcc gagctcggca tcgaactctc cgccttcgtc ctcttctcct 50100

ccgtcacagg cacctggggt acggcggggc aagccaacta cgcggctgcc aacgcctacc 50160

tggatgctct ggccgagcag cggcgcgccg acggcctcgc ggcgacgtcc atcgcgtggg 50220

gtccgtgggc cgagggcggc atggccgccg atgcggcact cgaagcccgt atgcgccgtg 50280

gcggagtacc gcccatgaag ggtgaggcag ccgtcaacgc ccttcagcgg gcgttgaacg 50340

cgaacgacac ggttgtcacc gtcgtggatg tggaatggga gcggttcgca cccggtttca 50400

ccgccgcacg ggcaagcacg ctcctcgccg aactgccaga ggcccagcgg gcacttgctc 50460

cgcaggaggg cgacgagggc caggacgacg gcgctgtcca cggtcgcggt ggtcactcgc 50520

ttgcggaacg gctcgcggag ctgtcggccg ccgagcgcga ccggctgctg ctcggcctcg 50580

tgcgcaagga agtcgccgcg gtactcggtc acgccggcgt ggaaagcatc ggtgcggcgc 50640

gcgcgttcaa ggaactcggc ttcgactcgc tcacggccgt cgaactgcgc aaccggctcg 50700

gcgcggtcac cgggcttcgg ctcccggcca cgctgatcta cgactacccc acgtccgggg 50760

ccttggcgga atacctgcgg ggcgagttgc tcggtacgca ggccgtggtg tccggtccgg 50820

tgtccaatgc cgtcgccgtc gacgacgacc cgatcgcgat cgtcgcgatg agctgccgct 50880

tccccggcgg cgtacggacc ccggaagacc tgtggcaact gctggcgacg ggacgcgacg 50940

ccatcggcga gttcccggaa gaccgtggct gggacgcgga ggccctgttc gggccccagt 51000

tcgagcagga cgccccgtat gcgcgtgagg gcgggttcct ctacgacgtc gccgacttcg 51060

atcccgcctt cttcgggatc tcgccgcgcg aggccctcgc catggacccg cagcagcgcc 51120

tgctgctcga aacctcctgg gaagccttcg agcgggccgg gatcgatccg ctctcggtgc 51180

ggggcagcca ggccggtgtc ttcgtcggca ccaacggcca ggactacctc tcgctcgtgc 51240

tgaactccgc ggacggcggc gacggcttca tgagcaccgg aaactcggcg agtgtcgtct 51300

ccggccgact ttcctatgtg ttcggcctgg aaggccccgc ggtcaccgtc gacaccgcgt 51360

gctcggcgtc cctggtcgcg ctgcatctcg cggtgcaggc gctgcgcaac ggcgaatgct 51420

ccctggcgct cgcgggcggt gtgacggtga tgtccacgcc cggcgccttc gccgagttca 51480

gccgtcagcg ggggctcgcg gaggacggcc gtatcaaggc gttcgcggcg gccgcggacg 51540

gtacgggctg gggcgagggc gtgggcatgc tcctggtgga gcggctctcc gacgcccgca 51600

ggaacggtca ccccgtcctg gccctggtcc ggggctcggc cgtcaaccag gacggcgcga 51660

gcaacgggct cacggctccg aacggcccct cgcagcagcg cgtcatccgt gccgctctcg 51720

cgagcgccgg cctggcaccc ggcgacatcg acgcggtcga ggcacacggc accggtacca 51780

agctcggcga cccgatcgag gcgcaggccc tgctcgccac gtacgggcag gaccgcccgg 51840

ccgaccggcc cctgcagctc ggttccatca agtccaacat cgggcacacg caggccgcgg 51900

ccggtgtcgc cggtttgatg aagatggtcc tcgccatgca gcacgggggtg ctgccgcaga 51960

ccctgcacgt ggacgagccg acccccacg tcgactggtc ggccggtgac atcgcgctgc 52020

tgaccgagcg gcgggagtgg ccggagacgg gccgtccgcg ccgggcgggc atctcctcgt 52080

tcggtgtgag cggtacgaac gcgcacacca tcctggagca ggcaccgccg ctcacggaga 52140

aggacgaggc tgaggccgcg aggccggaga ccggctccgc cgtctcggcg tggcccctcg 52200

cgggcaagac cgaagccggc ctgcgtgagc aggcggaacg gctgctggca cacatcgatg 52260

cccactccga gctgcggccg gtggacgtcg gtcactcgct cgcgaccggc cgggcggcgt 52320

tcgaccaccg tgccgtgctc gtggcgggag acgaccggtc ggagttccga cgggcactgg 52380

ccgcgctggc gtcgggagaa tccgtcgcgc aggtggtaca gggcatcgcg cgaccggatc 52440

agcaagtggc gttcctgttc acggggcagg ggagccagcg gctggggatg gggcgtgagc 52500

tgtacgagac gtatcccgtc ttcgcggatg cgctggacgc ggtgtgtgct cgccttgaac 52560

tgccgctgaa ggatgtgctg ttcggagggg acgcggatcg gctgaacgag accgcgtaca 52620

cccagccggc tctcttcgcg gtcgaggtgg cgttgttccg gctggtggag tcgtggggtg 52680

tgaggccgga cttcctggcc gggcattcga tcggtgagat cgcggccgcg catgtggcgg 52740

gggtgttctc gctggatgac gcctgtgctc tggtggaggc gcgtgggcgg ttgatgcagg 52800

cgctgccgac cggtggcgtg atgatcgcgg tccaggcgtc ggaggccgag gttctgccgc 52860

tgctgaccga gcgcgtgagc atcgccgcga tcaacggtcc gcagtcggtc gtgatcgcgg 52920

gtgacgaggc cgacgcggtc gcgatcgtgg acgcattcaa cgaccgcaag tccaagcggc 52980

tcgcggtcag tcacgcgttc cactcgccgc acatggacgg catgctcgcc gacttccgca 53040

aggtggcgga ggagctgtcg tacgaggctc cgcgcatccc catcgtctcg aacctcacgg 53100

gggccctggt caccgacgag atggggtcgg ccgacttctg ggtgcggcac gtccgcgagg 53160

ccgtccgctt cctggacggc atccgggccc ttgaggcgcc gggggtcacg gtgtacgtcg 53220

aactgggccc ggacggagtc ctgtcggcta tggcccagga gtgcgtcacc ggcgagggtg 53280

cggccttcgt gcccgctctc cgcaagggtc gtcccgaggc cgagacgatc acagcggccc 53340

tcgcccacgc gcacacccac ggcatcgccg tcgactggca ggcctacttc gccgggaccg 53400

gcgcccagcg cgtcgacctc ccgacctacg ccttccagcg ccagcgctac tgggtggatt 53460

ccttcgccga gttcgacgat gtcgcctcgg ccgggatcgg atcggccggt catccactgc 53520

tgggtgcggc ggtcgagctg ccggactcgg acgggttcct gttcaccggg cggctctccc 53580

tccgtacgca cccctggctc gccgatcacg tggtggcgga caccgttgtg gtgccgggcg 53640

cggcgttcgt cgagctggcg gtgcgcgccg gggacgaggt cggatgcgag gaagtggagg 53700

agctggttct tgaggcgccg ctcgtactgc ccgagaaggg ggccgtgcag ctgcggctca 53760

gcgtgggcgg ggcggacgac cagggacgcc ggtccgtaca cgtgcacagc cgcgttgagg 53820

cggccgatgg gggcggggtc cccggcgggg cgtggtcccg caatgcaacg ggtctcctct 53880

ccacccggcgg tagcggaagc gacgtcgact ccggcacggt catcggtgag tggccgccgg 53940

ccggagccga gcaggtggat gtgaccgcgg tacgcgaacg actggcggcc gcggggctcc 54000

accacgggcc gggcttccgg acgctgaccg aggtgtgggt gcggggcgag gaggtgttcg 54060

cggaggctag gctctccgac gaactgagcg cgtccgcagg gcggttcgcc ctgcacccga 54120

cgctgctcga cgccgcctcg caggcgctgg cggccggtac gaccgccgcc gcatccggca 54180

tcggtggtgc gggacggctg cctcaggcat ggcgcggggt acggctgcac gcggggggag 54240

cggacgctct gcgtctccgg atcaccgcgg gcggtcagga caccgtttcc gtcgtcctga 54300

ccgacacgca gggtgcgccg gtcgcgacgg tcggctcgct ggtcacggag gcggtcgacg 54360

ccgagcggta cgcggcggtt ccggacggat cccacgattc gctgttccgc ctcgactggg 54420

tgcggacgac ggctccgggg cggccgacct ccgcggactt cgcggtgctc ggtacccccg 54480

gcactggcat cggcgcccgc atcggcggtg acgagggctt cctcgtcggc gcgttggagc 54540

gggcgggtct gaccgccgag acgtacgacg gtctcgcggc gctcgactcg gccgtcgcgg 54600

ccgggatggc gatgccggaa acggtggtgg tgtcattcgc cgcagctttg gacccggcct 54660

cggactcggc cgcggacacg gtggcctccg tcgactcggc ggaggaggtc gcgcggctcg 54720

cccaggcggt gcgcgaggcg acgcaccggg cgctcgcgac cgtgcagggc tggctggaca 54780

acggccggtt cgccggagcg cgtctggtcg tcgtcacccg aggagcggtg gccacgggca 54840

gggacaccga ggtggaggac ctcgcccacg caccggtgtg gggtctgctg cgtgccgcac 54900

agaccgagca cccggaccgg ttcgtcctcg tcgacctcga cggggcggac gcctccgtcc 54960

gggccctgcc gggcgccatc gcctcgcagg agtccgaact ggccgtacgt gacggtgtgt 55020

tgtacgcgcc gcgcctggtc agggtcgggg cggaggcggt cacgggtgac accggcggtc 55080

gccgcatcga tccgcggggc acggtcctga tcaccggggc gagcggcgga ctcgccgggc 55140

tcttcgcccg ccatctggtg gcggagcacg gcgtacggca tctgctgctc accagccgca 55200

ggggcgccgc cgccgaaggt gccgcccaac tcgccgatga actcgtcgcg ttgggtgcgc 55260

aggtgacctg ggcggcgtgc gacgtggccg accgggacgc gctggccgca ctgctggcgt 55320

ccgtaccggc cgaacagccg ctgacggccg tcgtgcacac cgcggccgtc ctggacgacg 55380

gcgtcgtgga cctgctcacc cccgagcggg tggaccgggt gctgcggccc aaggcggaag 55440

cggcgctcca cctccacgag ctgaccaagg acctcgatct gtcggcgttc gtcctcttct 55500

ccgccgccgc cggcacgctc ggcggcgcgg ggcaggccaa ctacgccgcg gcgaacgtct 55560

tcctcgacgc cctcgcccgg caccgcacgg cccgtggtct caccgcgctg tccctcgtct 55620

ggggcatgtg ggccgaggag cggggcatgg cgggcaggct gacggaggcg gagctgggca 55680

gggcgggccg cggcggtgtg gcaccgctgt cggcgacgga ggggctcgcc ctcttcgacg 55740

cggccctcgc cgcggacgag gccgtgctcg taccggtcag gatcgatgtc ccgaccctgc 55800

gggcccgggc ggcggacggc gggatccacc cgatgttccg cggactggta cggactccgg 55860

tgcgcaggtc ggcgcagagc gcgggccgcg cggcgggcac cgtgcccacg gacggcgcgg 55920

gggagcggac gctggcccgg caactggccg agctgtccgt cgccgagcgg gagcggaccg 55980

tactggacct ggtacgcggc caggtggccg ccgtactcgg gtacgggtcc gccgaacaca 56040

tcggcggtga gcaggcgttc aaggaactcg gcttcgactc gctgaccgcg gtcgagctgc 56100

gcaaccgact cggcgcggcc ggcggtctga ggctgcccgc cacgctgatc tacgactacc 56160

cgaacccggc cgccctcgcc cagcacctgc tgagcgaggt ggccccggac acggcggagc 56220

gcaagctctc cgtactggag gaactcgacc ggctggagag caccttctcc tcgctggctc 56280

ccgcggaact gtccgcggcc gccggtgacg aggcggccca cgcgcgggtc gcggtacgcc 56340

tccagaccct gctggcccag tggaacgacg cccgtctggc agagggcggg agcggggccc 56400

acgcgatcga agaggcgagc gacgacgagc tgttcgccct catcgacaag aagttcggac 56460

agggctgaac ctcgcccacc gggcgcgccg ccgggtcagt ccccggcggc gccgcccacc 56520

cctgaaacga gacccgagac attccgagta cgtgcgaata ccgccacgat ctcggccacg 56580

cgaataggtg gaagcgccag tggcgaacga agcaaagctc cgcgagtacc tcaagaaagt 56640

cacgaccgat ctggacgagg cgtacggacg cctgcgggag atcgagagcc aggcccacga 56700

gcccattgcc atcacggcga tgagctgccg gttcccggga ggcgtacggt ctcccgaaga 56760

gctgtgggaa ctgctccgca ccggcgggga cgcactcacc gcgtttcccg cggaccgcgg 56820

ctgggacctc gacaacctgt tctcggacga ccccgacgac cacaacacgt cggtcacccg 56880

tgagggcggg ttcctcggcg aggcgtcctc gttcgacgcc gcgttcttcg ggatctcgcc 56940

gcgcgaggcc atggcgatgg acccgcagca gcggctgctg ctggagacct cgtgggaggc 57000

gttcgaacgg gccgggatcg acccccaggc gctgcgcggc agccagtccg gtgtgttcgt 57060

cgggatcaac gggtcggact acctgacccc gctgctggaa gcggccgagg actacgcggg 57120

gcacctgggg accggcaacg cctccagcgt gatgtcgggc aggctctcgt acacgttcgg 57180

cctggagggc ccggcggtca cggtcgacac ggcgtgctcc gcgtcgctgg tcgccctgca 57240

cctggccgtg caggcgctgc gggccggaga gtgctcgctg gccgtcgccg gcggggtgca 57300

cgtcatgtcc acgcccggac tcttcgtcga attcagcaag cagcgcggac tgtccacgga 57360

cggccgctgc aaggccttcg cggcgggcgc cgacggattc ggcccggcgg aaggcgtggg 57420

cgtcctgctg ctggagcggc tctccgacgc ccgcaagaac gggcgtccgg tccttgcggt 57480

ggtccgcggt tcggcggtca accaggacgg tgcgagcaac ggtctgacgg ctccgaacgg 57540

tccgtcgcag cagcgcgtca tccggcaggc cctcgccaac gcacggctct ccaccgacca 57600

ggtcgatgtc gtggaggcac acggcaccgg caccagcctc ggcgacccga tcgaggccca 57660

ggcgctcatc gccacgtacg gccaggaccg cccggccgat caaccgctgc tgctcgggtc 57720

ggtcaagtcc aacatcggtc acacccaggc ggccgccggt gtggccggcg tgatcaagat 57780

ggtgctggcg atgcagcacg gcgtgcttcc gcagagcctg cacatcgacg agccgtcgcc 57840

ccacgtggac tgggagtccg gcgcggtctc gctgctcacg gaacagacgg cctggcccga 57900

gacgacgcat ccgcgtcgtg cgggtgtgtc gtcgttcggg ttcagcggga cgaacgcgca 57960

tgtgatcgtc gagcaggctc cggtggttga ggaggtggcg ggggatccgg ccggtgtggt 58020

cgagggttcg ggtcccgggg tggtgccggt ggtgccttgg gtgttgtcgg gcaagagtgc 58080

gggggcgttg cgggcgcagg cggagcggtt gtccggattc ctcgcgggtg cttcggctgt 58140

ggatgtgccg tcggttgatg tggggtggtc gttggcgtcg tcgcgtgctg ggctggaaca 58200

ccgggctgtg gtgctgggcg atcacgcggc cggtgtggcg gcggtggcgt cgggtgtgat 58260

ggccgcgggt gtggtgacgg ggtcggttgt cggcgggaag accgcgttcg tgttcccggg 58320

gcagggctcg cagtgggtgg gtatggcggt ggggttgctg gattcctcgc cggtgttcgc 58380

tgcgcgggtg gaggagtgtg cgaaggcgtt ggagccgttc accgactggt cgttggtgga 58440

tgtgctgcgg ggtgtggagg gtgcgccgtc gttggagcgg gtggatgtgg tccagcccgc 58500

tctgttcgcg gtgatggtgt cgttggcgga ggtgtggcga gccgctggtg tgcgtcctgg 58560

cgcggtgatc ggtcattcgc agggtgagat cgctgccgcg tgtgtggcgg ggatcttgtc 58620

gcttgaggat gcggcgcggg tggttgcgtt gcgtagtcag gcgatcggcc gggtcctggc 58680

gggtctgggc gggatggtgt cggtgccgtt gccggcgaag gctgtgcggg agctgatcgc 58740

tccgtggggt gagggccgga tctcggtggc cgcggtgaac gggccgtcgt cggtggttgt 58800

ttcgggtgag gccgcggccc tggatgagct gctggtctcg tgcgagtcgg agggtgtgcg 58860

ggcgaagcgg atcgcggtgg attacgcgtc gcattcggct caggtggagt tgctgcggga 58920

agagcttgct gagctgctgg ctccgattgt tccgcgcgct gctgaggtgc cgttcttgtc 58980

gacggtcacc ggtgagtggg tgcgaggccc ggagctggat ggcgggtact ggttccagaa 59040

cctgcgtcgg acggtggagt tggaagaggc gacgcggacg ttgctggagc agggcttcgg 59100

tgtgttcgtc gagtcgagcc cgcacccggt gttgagcgtg ggcatgcagg agacggtcga 59160

ggacgcgggc cgggaggcgg ctgttctggg ctcgttgcgt cgtggtgagg ggggtctgga 59220

gcgtttctgg ctgtcgctgg gtgaggcctg ggtccgtggc gtgggtgtcg actggcatgc 59280

cgtgttcgcg ggcacgggtg cccagcgggt tgacctgccc acctacgcct tccagtcgca 59340

gcggttctgg ccggaggccg cgcccatcga ggctgtggcg gtgtcggcgg agagtgcgat 59400

cgatgcccgg ttctgggagg ccgtcgagcg cgaggacctg gaggcgctga ccgcggaact 59460

cgacatcgag ggcgaccagc cgctgaccgc actgctgccc gcgctgtcgt cgtggcgtcg 59520

gcagagccgt gagcattcga cagtggacgg ctggcgctac cgcgtcacct ggaagcggat 59580

cgctgagcct tccccggccc gcctgtcggg tacgtggctg gtcgtcgttc ccgaggtcgg 59640

cccggccgac gagtggacgg gagccgtcct gcgcatgctc gccgagcgcg gcgctgaggt 59700

ccgtaccgtg accgtcccgg ctgacggggc ggaccgtgac cggctcgccg tcacgctgaa 59760

ggccgagacg agcgaggtcg ctccgagcgg cgttctctcc ctcctcgccc tcgccgccgg 59820

tgcgggagcc ttcgccgccg aactcgccct gtgccaggcg ctcggtgacg ccgacgtggc 59880

cgcacctctg tggtgcgtga cgcgtggcgc tgtcgccacc ggccgttccg agcaggtggc 59940

cgaccccgcg caggcgctcg tctggggtct cgggcgggtc gcctccatgg agcaggggcg 60000

caggtgggga ggcctgctcg accttcccgc cgatctcgac ggccgtacgc tcgaacgtct 60060

cgcgggtgtc ctggccggtg atggttcgga ggaccaggtg gcgctgcgcg cctcgggtct 60120

cttcggtcgg cgtctggtgc acgcaccect cgccgacacc gccgccgtgc aggagtggcg 60180

tccgcagggc acgaccctgg tcacgggcgg tacgggcgcg ctgggcgcgc acgtggcccg 60240

ctggctcgcc gggaacggcg ccgagcacct gctgctcacc agccgacggg gccccgacgc 60300

gcccggagcc gccgcactcc gcgacgaact caccgccctc ggcacccagg tcaccatcgc 60360

gtcctgcgac atggccgacc gggacgccgt caccgccctc atcgccgcca tccccgccga 60420

ccagcccctc accgcggtga tccatgccgc ggcggtcgtg gacgacgggg tcatcgagac 60480

gctggccccg gagcaggtgg aggccgttct gcgggtcaag gtcgacgcga ccctcatcct 60540

ccacgagctg acccgtggcc tggacctgtc ggcgttcgtc ctcttctcct ccttcgccgc 60600

caccttcggc gcccccggcc agggcaacca ggcacccgga aacgcgtacc tggacgcctt 60660

cgccgagtac cgccgggggt cgggactgcc cgccacctcc atcgcctggg ggccgtgggg 60720

cagcgcggac ggcgacgaca gcgcggcggg cgaccggatg cgccgccacg gcatcatcgt 60780

gatgtcgccc gaacggaccc tcgtctccct ccagcacgcg ctggaccgtg acgagacgac 60840

cctgaccgtc gccgacatgg actggaagcg gttcaccctc gccttcaccg cggaccggga 60900

ccggccgctg ctcctggagc ttcccgaggc ccggcgcatc atcgagagcg cggagcggga 60960

gtccgccgac gacctggccg ggggagtgcc gctcacgcag cagctcgccg ggctgcccga 61020

ggtcgaacag gagcggctgc tcctcgacct ggtccgtacg gccgtcgccg ccgtcctcgg 61080

ccatgccgac ctggccgccg tcgaggcggg ccgggcgttc aaggagctcg gcttcgactc 61140

gctcacctcg gtcgaactgc gcaaccggct cggcgcggtc agcggtctga agctgcccgc 61200

cagcctggtc ttcgaccacc cgacccccgc cgccgtcgcg gccttcctac gcgccgggat 61260

cgtgcccgac gcggccgcgg gcggcgcgcc gctgctggag gagctcgaca agctcgaagc 61320

cgtactggag cggggcaccg ccgacaacgt cgtacgggcc cgggtgacca tgcggctcca 61380

gaagctcctg gggaagtgga acgagagcga ggaccagtcg ggcgccgagg tgtgggcggc 61440

cgcggccaac ggctccgggt cgggcatcgg cgcggggtcg gcggacggcg tgctggacga 61500

ggtcgagcag ctccaggagg cgagcgacga agagctgttc gccttcatca acaagggact 61560
cggccgcgcc tgaccgcaat ggatgtggat attgacggcg tgccgttaat tggccaggat 61620
agtcagcccc cttgttaatt tccacaaggc tcactgcccc ctgtcacacc ctcccaccca 61680
ggggtgtgta gggggcagtt aggggttgtc gggaagattg ggcggcgaat aacctgccgc 61740
tgagcagtcg attcaggcaa gaagtgaacc ggctgcatac ccgattcaat tctcggcttt 61800
atctgcacag ttattccgat gccgtctgct gcaaatgggt ggttgcgtta aatggcgaat 61860
gaagagacgc tgcgggacta cctgaagctg gtgacggcgg atctgcacca gacgcgacag 61920
cgtctgcgcg acgtcgaggc gaagaatcag gaccccatcg cgatcgtcgg catgggctgc 61980
cgctatcccg gcggtgtgac ctcgcccgag gagctgtggc agctcgtcgt ggacggtggg 62040
gacgccattt ccggcttccc cgccgaccgc ggctgggaca tggagacggt ctaccacccg 62100
gatcccgagc accccggcac gagctacgcc aaccagggtg gcttcgtccg ggacttcgcc 62160
cggttcgacc cgtcgctctt cggcatctcg ccgcgcgagg ccctcgccat ggacccgcag 62220
cagcggttgc tcctggagac ctcgtgggag gcgttcgagc gggccgggat cgacccgacg 62280
tcgatgcggg gcaagcaggt cggtgtcttc gtcggcacca gcaaccacga ctacctgtcg 62340
gcgctgctga gttcctcgga gaacgtggag ggctacctcg gcaccggcaa cgcggcgagc 62400
gtcgcctcgg gccggctctc gtacaccttc ggcctcgaag gcccggccgt caccgtcgac 62460
acggcctgct cgtcgtcctc ggtagccctg cacctggccg tgcaggcgct gcgcaacggc 62520
gagtgctcgc tcgccctcgc gggcggtgcc acgctgatgt cggctcccgg cacgttcatc 62580
gactacagca agcagcgcgg actggccacc gacggacgct gcaaggcgtt ctcgcccgac 62640
gccgacggct tcagcctcgc cgagggcgtg ggcatcctgc tggtcgagcg gctctccgac 62700
gcccgccgca agggacatcc cgtcctggcc gtggtccgtg gcaccgccgt caaccaggac 62760
ggcgccagca acggcctgac cgcgcccaac ggcccgtccc agcagcgcgt catccttcag 62820
gcgctgtcca acgccaggct caccccgac caggtcgacg cggtcgaggc ccacggcacg 62880
ggcaccggcc tcggtgaccc gatcgaggcg caggcgctca tcgccaccta cggccaggac 62940
cgccccgacg ggcggccgct gtggctgggt tcgctcaaga ccaacatcgg acacgcacag 63000
gccgcggccg gtgtcgcggg cgtcatcaag agcgtcatgg cgatgcgcca cggcgtgctg 63060
ccgcgcaccc tgcacgtgga cgagccgacc cccgaggtcg actggtcggc gggtgacgtc 63120
tccctgctca ccgaagcgcg gccctggccc ctgggcgacc agccgcgccg gatcggcgtc 63180
tcgtcgttcg gcatgagcgg caccaacgcc cacatcatcc tggagagcgc gcaggagtac 63240
gccgacggcc ggcaggccga cgccggtacc gcggggaacg aaccggccac cggccgtacg 63300
aacccgcccg gcgccctccc cgtcgtcctg tccggccgga ccgagcccgc cctgcgcgcc 63360

caggccgccg cgctgcacgc ccacctcgcg gcccaccccg gcctcggcat cgccgacctc 63420

gccttctccc aggccctcac ccgcgcagcg ctggaccggc gtgcggccgt cgtcgccgac 63480

gaccgcgacg ccctgctggc cgggctcgcg gcactggcgg aaggacgccc cagcgcggac 63540

gtggtcgaag gcagcgccac ggacggaaag ctggcgttcc tcttcaccgg gcaggggagc 63600

cagcggcccg gcatgggccg tgagctgtac gcgacgtatc ccgtcttcgc gcaggctctg 63660

gacgcggtgt gcgagcggct cgaactgccg ctcaaggacg tgctgttcgg gaccgacggc 63720

gccgccggcg ccgcgctcga cgagaccgcg tacacccagc ccgcgctgtt cgcggtcgag 63780

gtggccctct tccggctcgt ggagagctgg ggcctgaagc ccgactacct ggccgggcac 63840

tcgatcggtg agatcgcggc cgcgcacgtg gccggagtgt tctcgctgga ggacgcctgc 63900

accctggtcg aggcgcgtgg ccgtctgatg caggcgctgc cgaccggcgg cgtgatgatc 63960

gcggtcgagg cgtcggagga cgaggtcctg ccgctgctca ccgactgggt gagcatcgcc 64020

gccgtcaacg gcccccggtc ggtcgtcgtc gccggtgatg aggacgctgc ggtcgcgatc 64080

gcggaggcct tcgcagccca gggccgcaag accaagaagc tgacggtcag ccacgccttc 64140

cactcgccgc acatggacgg catgctcgac gccttccgca cggtcgccca gggactctcg 64200

tacgggactc ctcgcatccc ggtcgtctcg aacctcaccg gcgccctcgt caccgacgag 64260

atgggctcgg ccgacttctg ggtccggcac gtccgcgaag ccgtccgctt cctcgacggg 64320

atccgctggc tggagagccg cggggtcacc acctacatcg aactcggccc cggcggcgtc 64380

ctgtccgccc tcggccagga ctgccagacc gcgaccggcc cccgcgcggc cgccttcctc 64440

cccgcgctgc gcaccggccg ccccgaggcg tcgtcgctga ccgcggccgt ggccggcgcc 64500

catgtccgcg ggctctcccc ggactggacc gtccgcttcg ccggcaccgg cgcacagcgc 64560

gtcgagctgc ccacctacgc cttccagcgc gagctgtact ggccccgcga ccccttcacc 64620

gacccggccg aatccgccca cggcggcgaa ctcggcgcca ccgacgccaa gttctgggag 64680

gtcgtcgaca gcgaggacct cgccgcgctc gccgacaccc tcggggtcgg cggcgacgaa 64740

cccctcagca gcgtgctgcc cgcgctctcc gcctggcacc gccgccaccg cgaccgcgac 64800

accgtggacg gctggcgcta ccgcgtcacc tggaagccgc tgacggacac cacgcccgcg 64860

tccccctccg ggcactggct cctggtcgtc cccaccgagc acgccgacgc cccttgggcc 64920

gtcgccgccg agcgggcact gaccgcacgc ggtgtcaccg tgagcaccgt cgtgctcgac 64980

gcgaccctcg acgaccgggc cgccaccgcc cggcggatcg gcgaagccct cgctgcctcc 65040

gccgccaccg actccgcccc ggcgggcgcc gaaacgctcg ccggcgtgtt ctcgctgctc 65100

gccctggagg agcggccgca ccccgcggac ccggcactgt ccgccgggct cgccgccacg 65160

gtcgccctca tccaggcact cggcgacgcg ggagtggaag ccccgctgtg ggccgccacc 65220

tgcggcgcgg tctccaccgg ccgcaccgac cggctctcca gcaccgccca ggcgcaggtg 65280

tggggcctcg gccgcaccgc cgccctcgaa ctgcccgtgc gctggggcgg tctcgtcgac 65340

ctgcccggga cccccgacga gcgggccgcg ggccggctcg ccgacgtcct cggcggactc 65400

ggcggacccg gcgccgagga tcacctcgcc gtacgctcca ccggcgtctt cgtccgcagg 65460

ctggcccgcg ccacccgcga cgagcgcccc accaccgagt gggccaccac cggcacggct 65520

ctcatcaccg gcggcacggg cgcactcggc cgccacgtcg cccgctggct cgcccggacc 65580

ggggcgcagc acctgctcct ggtcagcagg cgcggcccgg aagccgaggg agccgacgcg 65640

ctcgccgccg aactgcgcgc actgggcgcc gaggtcacca tcgccgcctg cgacgtcgcc 65700

gaccgcgacg ccgtcgcggc cctgctcgcc accctcccgg ccgagcaccc gctgaccaac 65760

gtcgtgcacg ccgccggggt gctcgacgac ggcgtcctgg acgcccagac cccgcagcgc 65820

ctcgcggggg tcctgcgccc caaggcccac gcggcgcagg tcctgcacga gctgacccgc 65880

gacctggacc tctccgcctt cgtcctcttc tcgtccgtcg ccgccgtctt cggcgccgcc 65940

ggtcaggcca actacgctgc cgcgaacgcc tccttggagg ccctcgccga gcagcgccgc 66000

gccgacggcc tgcccgccac cgtgctggcc tggggcgcct gggccgaagg cggcatggcc 66060

accgacgaac tcgtcgccga gcgcctgcgg ctggccggac tgcccgccct cgcacccgaa 66120

ctcgccctgt ccgcactgca cagggcgctc accctggacg agaccgcctc gctcgtcgcc 66180

gacatcgact gggagcgcct ggcccccggc ctcaccgccg tacgcccctg cccgctgatc 66240

gccgacctcc ccgaggccgt gcacgccctc gccggagccg aggcgtccac cgggcccggc 66300

gccgccgccg acacgttcgc gcggcagctg gccgacgccc ccgccggtga acgcgaccag 66360

ctcgccctgg agttcgtacg cacccaggtc gcggccgtac tcggttacgc cggtcccgag 66420

tccgtcgacc cgggcagcgc cttccgggac ctcggcttcg actcgctcac cgcggtggag 66480

atccgcaacc tcctcacctc ccggaccggc ctgcgcctcc cggcgacgct gatcttcgac 66540

taccccaact ccctctccct ggccgccttc ctgcagggag aactgctcgg cgcgcaggcg 66600

accgaccccg cccgccacac ccccgcgggc cccggcaccg ccaccgatga cgaccccatc 66660

gcgatcgtcg cgatgagctg ccgcttcccc ggcggcgtac agagcccgga agacctctgg 66720

cagctgctct ccaccggccg tgacgcgatc tcgggcttcc ccggcgaccg cggctgggac 66780

ctcgacgggc tgtacgaccc cgagtccgcc ggggagaaca ccagttacgt ccgcgagggc 66840

ggcttcctcg ccggtgccac cgagttcgac cccgcgttct tcgggatctc cccgcgcgag 66900

gccctcgcca tggacccgca gcagcgcctg ctgctcgaaa cctcgtggga ggccttcgag 66960

cgcgccggaa tcgaccccgc caccgtgcgc ggcgaacaga tcggcgtctt caccggcacc 67020

aacggccagg actacctcaa cgtcatcctg gccgcacccg acggtgtcga ggggttcctg 67080

ggcacgggca acgcggcgag cgtggtctcc ggccgcgtct cctacgtcct cggcctggag 67140

ggcccggccg tcacggtcga cacggcctgc tcgtcctcgc tggtcgccct gcactgggcg 67200

atccaggccc tgcgccaggg cgagtgcacc atggccctgg ccggcggcgt gaccgtcatg 67260

tccacgcccg cctccttcat cgacttcagc cgtcagcgcg gcctcgcgga agacggccgt 67320

atcaaggcgt tcgccgcggc cgcggacggt acgggctggg gcgagggcgt cggcatcctc 67380

ctcgtcgaga ggctctccga cgcacagcgc aacggccatc cggtcctggc gatcgtgcgc 67440

ggctcggcca tcaaccagga cggcgccagc aacggcctca cggcgcccaa cggcccgtcc 67500

cagcagcgcg tcatccgcca ggccctcgcc agcggcggac tgacgacgat ggacgtcgac 67560

gccgtcgagg cccacggcac gggtacgaag ctcggcgacc cgatcgaggc gcaggcactc 67620

ctcgccacct acgggcagga ccggccggaa ggccgtccgc tgctcctcgg ctcgatcaag 67680

tcgaacctcg ggcacacgca ggccgccgcc ggtgtcgccg gtgtcatgaa gatggtcctc 67740

gccatgcagc acggtgtgct gccgcagacc ctgcacgtcg acgagccgac cccgcacgtg 67800

gactggtcgg cgggcgacgt cgccctgctg gccgatgccg tggcgtggcc cgagaccggg 67860

cgtccgcgcc gggcgggcgt ctcgtcgttc ggcatcagcg gcaccaacgc ccacaccatc 67920

atcgaacagg ccccggcagc cgtggcgccc gtcccgcccg tcgccaccac gcccgcacgg 67980

gccgacggac cgcagccgtg gctcctctcg gcgaagaccc gcgacgcact ccacgaccag 68040

gcgcgccgac tgcacgccca cgcggagctg aacccggaac tgagccccgc cgacctcgga 68100

ctctccctgg cggccggccg ttcggcgttc gagcggcgcg cggccgtgat cgccgcagac 68160

cgtgacgggc tgctggccgg cctcgcggcc ctggcggacg gcggcgcggc ggcaggactg 68220

gtggagggct caccggtcgc cggaaagctg gcgttcctgt tcaccgggca ggggagtcag 68280

cggctcggga tgggccgtga gctgtacgac acgtaccccg tcttcgcgga cgcgctcgac 68340

gcggtctgcg cgcatgtgga cgcgcacctc gaagtcccgc tgaaggacgt cctgttcggg 68400

gcggatacgg gtctgctgga ccagacggct tacacgcagc ccgcgttgtt cgcggttgag 68460

gtggcgttgt tccggctggt ggagagctgg ggtctgaggc ccgacttcct ggccggtcat 68520

tcgatcggtg agatcgcggc cgcgcatgtg gcgggcgtct tctcgcttca ggacgccagc 68580

gaactggtcg tcgcccgtgg gcggttgatg caggcgctgc cgaccggtgg cgtgatgatc 68640

gccgtccagg cgtcggagga cgaagtcctg ccgctgctga ccgaccgggt gagcattgcc 68700

gcgatcaacg gccctcagtc ggtcgtcatc gcgggtgacg aggccgacgc ggtcgcgatc 68760

gcggagtcgt tcacggggcg caagtccaag cgcctcacgg tcagccacgc gttccattcg 68820

ccgcacatgg acggcatgct ggaagacttc cgggccgtgg cggagggcct ctcgtacgag 68880

gctccgcgca tccccgtcgt ctcgaacctc accggcgctc tgatctcgga cgagatgggc 68940

tcggccgagt tctgggtccg gcacgtccgt gaggccgtcc gcttcctcga cggcatccgc 69000

acgctggaag ccgcaggcgt caccaagtac gtcgaactcg gccccgacgg cgtcctgtca 69060

gccatggccc aggactgcgt gagcggcgag ggctccgtct tcatccccgt actccgcaag 69120

gcgcgccccg agcccgagag cgtcaccacc gccctcacca cggcccacgt ccacggcatc 69180

cccgtcgact ggcaggcgtt cttcgccggg accggcgccc ggcgcgtcga cctccccacc 69240

tacgccttcc agcgccagcg ctactggccc gccgtctcct ccctctacct cggcgacgtc 69300

gaggcgatcg ggctcgacga caccgcgcac ccgctgctca gtgcgggtgt cgccctgccc 69360

gagtccgacg gcatggtgtt cgccgggcgg ctcgcgctct ccacccacgc ctggctcgcc 69420

gaccacgcca tcctcggcag cgtcctgctg cccggtacgg ccttcgtcga gctggccacc 69480

cgcgccggcg accaggtcgg ctgcgattac ctggaagagc tgaccctcga agcgcccctc 69540

gtcctgcccg agcacggcgg cgtccagctg cgcgtgtggg tcggcgccgc cgacgagtcc 69600

ggccgacggc cgttcgccct gcactcccgg gccgaaggcc tgccggtcga ggagccgtgg 69660

acgcggcacg ccggcggtgt actcgccgaa ggcgggcggc ccccggccga cttcgacctg 69720

acggcctggc ccccgccggg cgccgtcgaa gtggaccttg acgggcgcta cgaccagctc 69780

gacggcatcg gcttcgccta tggcccccacc ttccgtggcc tgcgtacggc ctggcagctc 69840

gacggcgaga tctacgccga ggtcaggctg cccgagggag ccgagggcga ggcgggccgg 69900

ttcggcctgc acccggccct gctcgacgcg gcactgcacg ccatcgggct gggcggcctc 69960

ggcgccgacg acggccaggg gaggctcccc ttcgcctgga gcggagtatc gctgcacgcg 70020

ggcggggctg ccgcactgcg cgtccacctc gctccggcgg gcgccgaggg cgtccgcctg 70080

gagatcgcgg acgcctcggg cgcaccggtc gcggccgtcg agtcgctcgg gctgcgcccg 70140

gtgacggccg agcagctccg tgccgctcgt gccacctacc acgagtccgt gttccgtcag 70200

cagtggaccg agctgccggg tctcggcgct ccggccgcga cccccgccgt ccggtacgcg 70260

ttcctcggcg gcgacagcgg cgacagcggc gacagcggtg acaccgcagc cgccgaccgt 70320

caccaggacc tggcggcgct cgccgccgcg atcgacgccg gaaggcccgt accggacgag 70380

gtggtcgtcg aactcgccgc cgcgccctgg gccgtgtcgg cgtcggccgt gcacagtgcc 70440

gcgcacgatg cgctggcact catccagacc tggctcgcgg acgaccggtt cgccgccgca 70500

cgcctggtgt tcctcacccg cggcgcggtg gccgcggacg cgggcgacga cgtgaccgat 70560

ctcgccgccg ccaccgtgtg gggcctgctg cggtccgcgc agacggagaa ccccggcagg 70620

atcgccctcg tcgacaccga cggccacgac cggagcgagc aggccctgcg ggcggcgctc 70680

acctccgacg aggagcggtt cgcgctgcgc gccggagcgg tcctcgtgcc ccggctcgcc 70740

cgggtcgaga tccagcagga cgactccgcc cggacaccgg ccctcacgcc cggcggcacg 70800

gtactgatca ccggagccac cggagcgctg ggcggtctct tcgcccggca cctcgccgcc 70860

gaacacggcg tggagcggct gctcctcgtc ggcaggcgcg gggccgacgc ccccggcgcg 70920

gccgaactcg tcgccgaact cgccgagtcg ggcaccctcg ccacctgggc ggcgtgcgac 70980

gtggccgacc gggacgcgct cgcggcactg ctcgcggaca ttcccgccga gcacccgctg 71040

accgccgtcg tccacacggc cggagtcctc gacgacggcg tcatctcctc gctgacgccc 71100

gagcggctct ccgccgtgct gcggcccaag gtggacgcgg cctggaacct gcacgagctg 71160

acccggggcc tcgacctcgc cgccttcgtg ctcttctcct ccacctccgg cctcttcggc 71220

ggccccggac agggcaacta cgccgccgcc aactccttcc tggacgccct cgcccagcac 71280

cgccgcgctc acgggctccc cgcgacctcg acggcctggg gcctgtggtc cgtggccgac 71340

ggcatggcgg gcgccctgga cgcggccgac gtcaaccgca tgcggcgggc cggactgccg 71400

ccgctgaccg ccgccgacgg cctcggcctg ttcgacacgg cggtctccct cgacgaggcc 71460

tccctggccc tgatgcgggt ggacaccgaa gtcctgcgca cccaggccgg ggccggtacc 71520

atcgcgccgc tgctgcgcgg tctcgtacgg ggcgtggccc gccggtcggt cgacgtgtcg 71580

gccggtgccg ggggcgccga atcggagctg cgcggcaggc tggcggcgct caccgccgcc 71640

gagcaggacc gggcgctgct ggacctggtg cgtacgcagg tcgcggcggt cctcggacac 71700

gccggacccg cggccgtgga gtcgggacgg gccttcaagg aactcggttt cgactcgctc 71760

accgcggtgg agctgcgcaa ccggctgaac gccgccaccg cgctgcgcct gcccgcgacg 71820

ctgatcttcg actatccgga cccgaccgtt ctcgcccggt acctgcgcgg cgagctgatc 71880

ggtgacgaca ccacggacgc cgtggccgag ccgctcacgg ccgtggccga cgacgagccc 71940

atcgccatcg tcgccatgag ctgccgctac cccggtgacg tacgcacccc cgaggacctg 72000

tggcagctgc tgacggcggg cgccgacggc atcacccggc tccccgagaa ccggggctgg 72060

gacaccgagg gcctgtacga cccggacccg gagagccagg gcacctcgta cgcccgcgac 72120

ggcggattcc tgcacgacgc ggccgagttc gacgcctcct tcttcgggat ctcgccgcgc 72180

gaggccctcg ccatggaccc gcagcagcgc ctcctcctgg agacgacctg ggaggtcttc 72240

gaacgggccg gcatcgcgcc gtccgcggtg cgcggcagcc ggacgggtgt cttcgcgggt 72300

gtcatgtacc acgactacgg cgcgcgcctg cacgccgtgc ccgacggcgt cgagggctac 72360

ctcggcaccg gcagctccag cagcatcgtg tcgggccggg tcgcctacac cttcggcctg 72420

gagggcccgg cggtcaccgt cgacacggcc tgctcctcgt cgctggtcgc cctgcacctc 72480

gcggcccagg cgctgcgcaa cggcgagtgc tcgctcgctc tcgcgggcgg tgtcaccgtg 72540

atgttcacgc ccggaacctt catcgagttc agccgtcagc gcggcctggc cgccgacgga 72600

cgctgcaagt ccttcgcggc cgccgccgac ggcacgggct ggggcgaggg cgcgggcatg 72660

ctcctgctgg agcggctctc cgacgcgcga cgcaacggcc accaggtcct cgcggtcgtc 72720

cgcggctcgg ccgtcaacca ggacggcgcc agcaacggcc tcaccgcccc gaacggcccc 72780

tcgcagcagc gcgtcatccg gcaggccctc gccaacgccg gtgtcgccgc cggacacgtc 72840

gacgccgtcg aggcacacgg caccggcacc accctcggtg accccatcga ggcgcaggcc 72900

ctgctcgcga cctacggcca ggagcacacc gacgaccggc cgctgctcct cggctcggtg 72960

aagtccaacc tcggtcacac acaggccgct tcgggcgtcg ccggtgtcat caagatggtc 73020

atgtcgatgc ggcacggtgt gctgccgaag accctgcacg tcgacgagcc gaccccgcac 73080

gtggactggt cggcgggcgc ggtctcgctc ctcaccgagc agaccccgtg gcccgagacc 73140

ggccgtccgc gccgcgcggg cgtctcctcc ttcggcatca gcggcaccaa cgcgcacgcc 73200

atcatcgagc aggccccgga gccggacccg gcccgggcga aggcgacggc gcggcccgcg 73260

ccggacgccg cggcgccgtc gtccgtgccc ctgatcgtgt ccgcccgcgg cgaggacgcg 73320

ctgcgcgccc aggcccgcag gctccacgcc cacgtccacg ccgaccccgg cctgcgcgcc 73380

gtcgacctcg gcctctccct ggcgaccacc cgctcggccc tggagcagcg cgcggcgctg 73440

gtggccggcg accgcgcgga actgctgcgc ggcctggacg ccctggcccg cggcgaggac 73500

accgcggggc tggtgcgcgg caccgcccgc gagggccagg tggcgttcct gttcaccggt 73560

cagggcagcc agcggccggg gatgggacgc gagctgtacg acgcgcatcc cgtcttcgcg 73620

gacgcgctcg acgagatctg cggcgaactg gaccggcacc tcgaagtacc gctcaagggc 73680

gtgctgttcg cgaccgaggg cgatctgatc caccagaccg cgtacacgca gcccgcgctg 73740

ttcgccgtgg aggtggccct gttccggctc ctggagagcc ggggcgtgca gcccgacttc 73800

ctggccggtc actcgatcgg tgagatcgcc gcagcccatg tggcgggcgt cttctcgctc 73860

caggacgcca gtgaactggt cgccgcccgt gggcggttga tgcaggcgct gccgaccggt 73920

ggcgtgatga tcgccgtcca ggcatcggag gacgaggtcc tgccgctgct gacggaccgg 73980

gtgagcatcg ccgcgatcaa cggcccccag tcggtcgtga tcgcgggcga cgaggccgac 74040

gcggtggcca tcgccgagtc cttcacggac cgcaagtcca agcggctcac ggtcagtcac 74100

gccttccact cgccgcacat ggacggcatg ctcgccgact tccgcaaggt cgccgagggc 74160

ctcgtctacg agaacccgcg catcccggtc gtctcgaacc tcacgggggc cctggtcacc 74220

gacgagatgg gttcggccga cttctgggtc cggcacgtcc gcgaggccgt ccgcttcctc 74280

gacggcatcc gcgccctgga agccgcgggc gtcaccacac acatcgagct gggccccgac 74340

ggcgtgctct gcgccatggc ccaggaatgc gtgagcggcg aggacaccgt cttcgtcccc 74400

gtactgcgcc ccggccgccc cgaggccgag accgtcacca ccgccctcgc ccgcgtccac 74460

gtccagggcg tacccgtgga ctggcaggcg tacttctccg gcaccggcgc ccagcgcgtc 74520

gacctgccca cctacgcctt ccagcgcaag cgctactggc tcgacgtcgg cgtctccgtc 74580

gaggacgtgc tggcggccgg tctcgatgcg gccgaccacc ccctgctggg cgccaccgtc 74640

tccctgcccg gatccgacgg gctggtcctc accggacgcc tcgcgctgtc cacgcacccc 74700

tggctgagcg accacaccgt catggacacc gtcctgctgc ccggcacggc cttcgtcgaa 74760

ctcgccctgc gggccggtga actggtcggc tgcggcgccg tcgaagagct ggcgctcgaa 74820

gccccgctca ccctcgccga ccagggcgcc gtccagttcc agctggccgt ggacgcgccg 74880

gacggcgccg ggcgccggac cctgaccctg cactcccgcc gcgcgggtgc cccggccgaa 74940

gagccgtgga cacggcacgc caccggcgtt ctcacgcccg aagcgtccgc cgtgcccgcg 75000

caccccttcg acctgaccgc atggccgccg gccgacgcgg agcccgtgcc caccgacgcc 75060

ttctaccccg gcgcggccgc ggccggcctc ggctacggac cggtcttcca ggggctgcgg 75120

gccgcctggc ggcgcggcga cgaactgttc gccgaggtcg cactcgacga ggagcacgag 75180

gccgacgccg ccgcctacgg gctgcacccc gccctgctcg acgcggccct gcacgccatc 75240

ggcctcggag cgcccggcgc gcccgccgac gccccggccg aaggagcccg gctgccttc 75300

gcctggaccg gcgtacgcct gtacgcggcc ggcgcggcgg gcatccgcgt ccggctgacc 75360

gccgccgcat ccggcggcat cgccctggac gtggccgact ccaccggagc gccggtggcc 75420

tccgtcgagt ccctgatcct gcgccccgtc tccgcggagc agctcggcgg ggaccgcacg 75480

gcccaccacg agtcgctctt cggcgtcgag tggaccaggc tgtcctctcc caccggtgcg 75540

atcccctccg gcgaacgctg ggccgtactc ggcgaggacg agccggacct ccgggtcggc 75600

ggcgaacgcc tcgacgtgta cagcggtctc acggcgctgc gcgaggaaat cgccgcgggc 75660

acctcggcgc cggacgtcgt cgtcgtaccc ctgtcctccg ccgcgtccgg tggcggacgt 75720

gcggggaccg cccgggccgc cgcgcaccac gcgctggccc tggtcaagga gtggctggcc 75780

gacgaacggc tcgacggcgc acggctcgtg ctgctgaccc ggggcgcggt ggccgccgta 75840

cccgacgagc acgtgaccga tctgacccac gcccggtgt ggggcctcgt acggtccgcg 75900

cagtcggaga accccggccg gttcgtgctc gccgacaccg acggcgccga cgcctccttc 75960

ggggcgctgg ccgccgcgct cgccaccgac gagccgcagc tcgccctgcg gtccggcgag 76020

gcacacgcct tccggctgcg ccgcatcgcc cgtaccgcga gcgatccggc cggtgaaacc 76080

ggcacgggcg acggccccac ccgtgccgac gacgccggga ggatcgccgc cgacggcacg 76140

gtcctggtca ccggcgcgag cggcaccctc ggcgggctct tcgcccgcca cctggccacc 76200

acgcacggcg cacggcacct gctgctgctg agccgtcgcg gggaccgggc ccccggggcc 76260

ggggaactga cccgtgagct gaccgaagcg ggcgtggacg tgacctgggc ggcgtgcgac 76320

gcggccgacc gggacgcgct cgccgccgta ctcgccgcga tcccggccga ccggccgctg 76380

acggcggtcg tccacaccgc cggtgtgctc gacgacggca tcatcgactc cctcacaccc 76440

gaacgcctcg acaccgtgct gcggcccaag gtcgacgcgg cctggaacct gcacgagctg 76500

accgagggcc acgaactctc cgccttcgtg ctcttctcct cggtcgccgg ctgcttcggc 76560

gccgcgggcc agggcaacta cgcggcggcc aacaccttcc tggacgccct cgcccagcac 76620

cgcaaggccc ggggcctcac cgccagttcc ctcgcctggg gcctgtggga gacgacggac 76680

ggcatggccg gcgcgctcga cgaagccgac ctgacccgca tggcccgctc cggtgtggcc 76740

gcgctcgccc ccgacgaggg cctggccctc ttcgacacct cccgcaccct ggacgacgcg 76800

gtcctcgtcc ccatgcggat cgaactgggc gcgctgcgcg cccaggccgc ggacggcacc 76860

ctgccgccgc tgctgcgcgg actggtgcgc actcccgcgc gccgggccgc cggctccacg 76920

gcacgcgccg gaacgcgccc cggcaccgac ccggcgggca ccctcgaaga gcgcctcgcc 76980

ggactgtcgg ccgccgaacg cgaccgggcc ctcatggagc tggtccgcac acaggtggcc 77040

gcggtcctgg gctacgcggg ccccgacgac gtcgacgccg cacggggctt cctcgacctg 77100

ggcttcgact cgctcacggc cgtcgacctg cgcaaccgcc tcacggcgag cgccggactc 77160

cggctgcccg tcacgctcat cttcgactac ccgtctccga ccgcgctcgc cgcgtacctc 77220

gccgaacgcc tcggccaggg cgacccgtcc cgccggcccg tccacgcgga actcgacaag 77280

ctcgaatcga tcctctcgac ggtcggcccc gacgacgtcg aacgcgcggg catcaccgcc 77340

cggctgcgag accttctggc gaagtggaat gaaacgcaca gtgcacagga cagcgccgca 77400

gacgagcggg aaatccagtc cgcgacggcc gacgagatct tcgatctcct cgacgacgaa 77460

ctcgggctgt cctgaccggc tcctgcccgg cgggcggccg gccggtgcgg agcaccggct 77520

cccggccgcc cgcccgtccg gcacccacct tccgatccac cggctccgcg cgagctttcc 77580

gactctgacc acggggatgg cgtaaatggt gaacgaggag aagtacctcg attacctcaa 77640

gcgggcgact accgacctcc gcgaggcacg acgacggctg cgcgaggtgg aggaacggga 77700

gcaggagccg atcgccgtcg tggcgatgag ctgccgctac cccgggggga tcgacacccc 77760

cgagaagctg tgggacctcg tcgcccacgg ccgggacgcc gtctccgcct accccacgga 77820

ccgcggctgg gacgccgaag tcctcttcga ccccgacccc gagaccggga tcgaggcgta 77880

cgaacaggtc ggcggcttcc tgcacgacgc ggccgacttc gaccccgcgt tcttcgggat 77940

ctcgccgcgc gaagccctcg ccatggaccc ccagcagcgg ctgctgctgg aaacctcctg 78000

ggaggcgttc gagcgggccg gaatcgaccc ggcgaccctg cgcggcagcc gtacgggcgt 78060

cttcgccggc ctgatgtacc acgactacgc cgcccggctg ttcagcgtgc ccgaggagat 78120

cgagggcttc ctcggcaacg gcagctccgg cagcatcgcc tcgggccgga tcgcctacac 78180

cctcggcctc gaaggccccg ccgtcaccgt cgacacggcc tgctcctcct cactggtcgc 78240

cgtgcacctc gcggcccagg cactgcgcaa cggcgagtgc acgctcgccc tcgccggtgg 78300

tgtcaccgtc atgtcgaccc ccggcacctt caccgagttc agccgccagc gcggcctggc 78360

ggccgacggc cgctgcaagt ccttcgcggc cgcggcggac ggtacgggct ggggcgaagg 78420

cgccggcatg ctcgtcctgg aacggctctc cgaagcccgc aggaacggcc accccgtcct 78480

ggcactcgtg cgcggttcgg ccgtcaacca ggacggcgcc agcagcggtc tgacggcccc 78540

caacgggccg tcccagcagc gcgtcatccg ccaggcactc gccggtgcgc ggctgtcggc 78600

cacccaggtc gacgcggtcg aggcccacgg caccggcacc accctcggcg acccgatcga 78660

agcgcaggcc ctgctcgcca cctacggcca ggaccgtccc gacggccgcc cgctgtggct 78720

gggctccatc aaatcgaaca tgggtcacac ccaggccgcc gccggtatcg cgggcattat 78780

caagatggtc atggcgatgc gccacggcat cctccccaag accctgcacg tcgacgagcc 78840

gaccccgaac gtcgactggt ccgagggcgc ggtctccctg ctcaccgagt ccgtgccgtg 78900

gcccgagacc ggcgcgcccc gccgcgcggg agtctcgtcg ttcggcatca gcggcaccaa 78960

cgcccacacc atcctcgaac aggccccgga cgccgtcgag gccgcacccg ggaccgagcc 79020

ccccgcggcg gccgcaccgc ccgtgccccc gctctggacc ctctccgcca agagcccggc 79080

cgcgctgcgc gcccaggccg ggaaactgca cgcccacctg accgcacacc ccggcctgcg 79140

ccccggggac atcgcccact cgctcgccgt cggacgcacc gacttcgagc accgcgccgt 79200

cctcacctcc gccgacgggc ccgtgggcct cgtccgtgcg ctggaagccc tcgcggactc 79260

ggctcccgag gacacggcac ccgccgacag ggcaccgggg gtcacccggg gccgcccggt 79320

cgccgggaag ctggcgttcc tgttcaccgg gcaggggagc cagcggctgg ggatgggccg 79380

cgagctgtac gagacgtatc ccgtcttcgc gcaggctttg gacgcggtgt gtgagcggct 79440

gaatctcgaa gtgccgctga gggatgtcct gttcggggcg gatgcgggtc tgctggacca 79500

gacggtctac acgcagaccg cgttgttcgc ggtcgaggtg gcgttgttcc ggctggtgga 79560

gagctggggt ctgaagcccg acttcctggc gggtcattcg atcggtgaga tcgcggccgc 79620

gcatgtggcg ggggtgttct cgctggagga tgcgtgcgcg ctggtgtcgg cgcgtggccg 79680

cttgatgggt gcgctgccgg gtggcggcgt gatgatcgcc gtccaggcgt cggaggacga 79740

ggtcctgccg ctgctcaccg accgcgtgag cattgccgcg atcaacggtc cgcagtcggt 79800

cgtgatcgcg ggcgacgagg ccgacgcggt ggcgatcgcc gagtccttcg cggaccgcaa 79860

gtccaagcgg ctcacggtca gtcacgcctt ccattcgccg cacatggacg ccatgctgga 79920

ggacttccgg gccgtggcgg agggcctgtc gtacgaggcc ccgcgcatcc ccgtcgtctc 79980

caacctcacc ggcgccctcg tctccgacga gatgggctcg gccgacttct gggtccgcca 80040

cgtccgcgag accgtccgct tcctcgacgg catccgcgcc ctcaccgagc gcaacgtcgt 80100

ccacttcgtc gaactcggcc cggacgccgt gctgtcggcc atggcccagg actgcccctc 80160

cgccgacacc gcggccttcg tgcccgtact ccgcaagggc cgttcggaga ccggttcgct 80220

gaccgacgcc ctcgcgcggc tccatgtggg cggggtggcc gtcgactggg acgcgtacta 80280

ctccggtacg gacgtccagc gcgtcgacct gcccacctac gccttccagc gcgcgcacta 80340

ctggctcgac gcaggccggc ccctcggcga cgtctcctcg gccgggctcg gtgcggccgg 80400

ccacccgctg ctcggggccg ccgtggccct cgccgacctc gacggtttcc tctacaccgg 80460

ccgtctctcg ctcgacaccc accctggct cgccgaccac gccgtcatgg gttcggccgt 80520

actgccgggc accgccttcg tcgaactggc catccgcgcc ggtgaccagg tcggctgcga 80580

cctgctcgaa gaactcaccc tgcacgcacc gctcgtactg cccccggccg gaggtgtgca 80640

ggtccagttg tgggtcggcg caccggacgc caccggccgc cgcaccctgg gtgtgcactc 80700

ccgccccgag cccgcaccgg acgccgtcgg cccggacgcc gacgcggcgg agccgtggac 80760

ccggcacgcc gacggtgtgc tcgccacggg tgccccgcag ccgtccttcg cccccgacgt 80820

ctggccgccg gccggtgcca ggcccctgcc cgtcgacgag ctgtacgccg ggctcgccga 80880

ggcgggcctc gaatacggcc ccgccttcca gggcgtccgc gcggcctggg cgagcgacga 80940

cgcggcctac gtcgagatcg cggccgccga cggacagtgg gccgatgccc cgctgttcgg 81000

actgcatccc gcgctcctcg actcggcgct gcacgccatc ggtctggccg ggctcgtcga 81060

ggacaccggc cgcggccggc tgcccttctc ctggtccggg gtgtccctgt acgccgtggg 81120

cgcctcggtg ctgcgcgtac ggctggccaa ggccggaccg gacgcggtgt ccctggccct 81180

cgccgacggc gccggacagc ccgtgggcga catcgcctcg ctcaccctgc gccctgtctc 81240

ggccgagcag ctggacaccg ggcggggcgg tcaccatgac gcgctgttcc aggtggactg 81300

gaccccgctg aacctgcccc gtgctgtcga cagccgctgg gccgtgctcg gcgagcccgt 81360

ccccaccgac gagccgggcg acggcgtggc gcgccacgcg gacgcggagg cgctgagcgc 81420

ggccctcgac gcgggtgctc cggtgccgga tgccgtactc gtacgccacc ccgccctgcc 81480

cgaacccacc cccgaggcgg tccaccaggc cgcgcaccgg accctcggcc tgctgcggca 81540

ctggctcggc gacgaccggc tcgccgacag ccgcctcgtc ctgctcacgc acggcgcggt 81600

cgccgcggga gacgcggacc aggtacccga cccggtgcac gccgtggtct gggggctggt 81660

ccgctccgca cagtccgagc acccgggccg gttcctgctg atcgacagcg attccggtat 81720

cgacacactc tcctggccga cgttcggtgc cgttctcgcc tccgaggagc cgcaggtcgc 81780

cctgcgcggc ggcgtggccc acgcacccag gctggccaag gttcccgcca ccgctaccgc 81840

cgctgccgtc gtcgagacgt cgtcgtacga ccctgacggc accgtcctcg tcaccggggc 81900

cagcggcacg ctcggcggac tcgtcgcccg tcacctcgtg accgggcgcg gcgtacggcg 81960

tctgctgctg ctgagccgtc ggggcgccga tgccccggt gccggtgaac tggccgctga 82020

gctgaccggg ttgggtgccg aggtgtcgtg ggcggcgtgt gacgcgggtg accgcgacgc 82080

gctcgcggcc gtactggccg ccgttcccgc agcgcacccg ctcaccgcgg tcgtccacac 82140

ggccggtgtc ctcgacgacg gcgtgatcgg ttcgctcacc ccggagcgcc tcgacacggt 82200

ccttcgcccg aaggccgacg ccgctctcca cctgcacgaa ctgacccgcg acctgcccct 82260

gaccgccttc gtcctcttct cctccgcggc cggggtcttc ggcgcaccgg gtcagggcaa 82320

ctacgccgcc gccaactcct tcctggacgc cctcgcccag taccggcgtg cccacgggct 82380

ccccggccgg tcgctggcct ggggcctctg ggaggacgcc gaaggcatgg cgggcgccct 82440

cgaccgcgcc gacctcgacc ggatgaagcg cggcggagtc cacggactca ccgcctccga 82500

gggcctcgcg ctcctcgacc tcgccgacgc cctcggcgcg gaccgtgacg accagggcca 82560

ggatcaggag acggccggac gggcgctgct cgtgccgatg cggctgaccc ttcccgccgt 82620

cgcccccggc gccgaagtcg ccccgctgtt ccggggattg gtccgcaccc ccgcgagacg 82680

cgtcgcggcc ggagccacca cgggagccac caccggaacc gggcccgacc tctccgctct 82740

cgaacggcgg ctcctcggcc tcgacgcgcc ggagcgggag cggctgctcc tcgacctcgt 82800

ccgcggccat gtcgccgacg tgctcggcca cggctccccg gacgccatcg accccgaaca 82860

ggccttcagc gagctgggct tcgactccct gacggcggtg gaactgcgca accgcctggg 82920

cgcggccatc ggccggcggc tgcccgccac gctgatcttc gaccacccgg cctcgctcac 82980

cctcgcccgt cacctctccg gtgaactcgc cgggtcccag gccgcgttgg cgccagccgg 83040

gcccgcgccc accgtgaccg acgacgaccc gatcgccatc gtggcgatga gctgccgcta 83100

ccccggcggc gtgaccaccc ccgaggagct gtggcagctc ctcgcgggcg gcggggacgc 83160

gatatccggc ttccccgccg accgcggctg ggacgtcgag tcgctgtacg accccgatcc 83220

cgaccacccg ggcacctcgt acacccgcca cggcggcttc ctgcgcgacg ccgccgcgtt 83280

cgatccgacg ttcttcggga tcagcccgcg cgaggccgtc gggacggacc cgcagcagcg 83340

gctcctcctg gagaccacct gggaggcgtt cgaacgggcc gggatcgacc cggccaccgt 83400

gcgcggcagc cggaccggtg tgttcgcggg cgtcatgtac cacgactacg cggccctgct 83460

ggagcgctcg aaggacggag cggacggctc cctcggctcg ggcagcaccg gcagcatcgc 83520

ctcgggccgg gtctcgtaca ccttcggtct cgaaggcccc gccgtcacga tcgacaccgc 83580

ctgctcgtcg tcgctcgtgg ccctgcacat ggccatccag gcgctgcgca ccggcgagtg 83640

cgacatggcg ctggccggcg gtgtcaccgt catggcgacc cccggcacgt tcatcggctt 83700

cagccgtcag cgcggcctgt ccgccgacgg ccgctgccgc gccttctcgg ccgacgccga 83760

cggtacgggc tggggcgagg gcgtcggcat gctcctcgtg gaacgcctgt ccgacgcccg 83820

ccgcaacggg catccggtcc tggccgtggt ccgtggctcg gcgatcaacc aggacggcgc 83880

gagcaacggc ctcaccgccc ccaacggccc ctcgcagcag cgcgtgatcc gcgcggccct 83940

cgcgagcgcg ggcctgtcgg ccgccgaggt cgacgcggtc gaggcgcacg gcaccggtac 84000

gacgctcggc gatccgatcg aggcgcaggc gctcctggcc acctacggcc gggagcacac 84060

cgaggacagc ccgctgtggc tcggctcgat caagtccaac atgggtcaca cgcaggcggc 84120

cgccggtgtc gcgggcgtca tcaagatggt cctcgccatc cagcacggcg tgctgccgcg 84180

caccctgcac gcggaccggc cctcgcccca cgtggactgg tcgcagggcg ccgtctcgct 84240

gctcaccgag tccgtcccgt ggccggagac gggccgtccg cgccgcgcgg gcgtgtcgtc 84300

gttcggcatc agcggcacca acgcgcacac gatcatcgag caggcgccgg aggaggccac 84360

ggtggccccg gccgacgcgg tggccgcgcc gagcgcgctg cccctgcagc tcgcgggccg 84420

cagcgccgag gcgctctccg cccaggcccg tgcgctgagc gcacacctga ccgcacaccc 84480

cgacgtcccc ctcgcagacc tcgcctactc cctggccacg agccgtgcca ccttcgacca 84540

ccgggcggtc ctggtcgcga cggagggcac aacggccgcc acggccgtca cggcgctcga 84600

cgccctcgcc gaccggcgca cggcaccggg cctggtgcgg ggcacggcca gcaagggcgg 84660

tcgcacggcg ttcctgttca cggggcaggg gagccagcgg ctggggatgg ggcgtgagct 84720

gtacgaggcg catcccgtct tcgcgcgggc tctcgacgcg gtgtgtgatc gcctggaact 84780

gccgctgaag gatgtgctgt tcggtactga cgcgggtctg ctgaacgaga ccgtgtacac 84840

gcagccgggt ctcttcgccg tcgaggtggc gctgttccgt ctgctggaga gctggggtgt 84900

gaagcccgac ttcctggccg ggcactcgat cggtgagatc gccgcagccc atgtggccgg 84960

ggtgctctcc ctcgatgacg tgtgcgctct ggtggaggcg cgtgggcggt tgatgggtgc 85020

gctgccgggc ggtggcgtga tgatcgccgt ccaggcgtct gaggctgagg tcctgccgct 85080

gctgaccgac cgggtgagca ttgccgcgat caacggcccc cggtcggtcg tcatcgcggg 85140

cgacgaggcc gacgcggtcg cgatcgtgga gtccttcacg gaccgcaagt cgaagcggct 85200

cacggtcagt cacgccttcc actcgccgca catggacggc atgctcgacg ccttccgtga 85260

aatcgcggag ggtctgtcgt acgaggctcc gcgcatcccg gtcgtctcca acctcaccgg 85320

ggccctggtc tcggatgaga tgggttcggc ggacttctgg gtgcggcacg tccgtgaggc 85380

cgttcgtttc ctggatggca tccacgccct ggaggccgcg ggcgtgacga cgtacgtcga 85440

actcggcccc gacggagtcc tgtcggcgat ggctcaggag tgcgtgaccg gcgaggactc 85500

cgtcttcgtg ccggtcctgc gctcgggtcg tcccgaggcc gagagcgtca ccacggccct 85560

cgcccaggcg catgtccgcg ggatcgccgt cgactggcag gcgtacttcg ccgggaccag 85620

tgcccagcgc gtcgacctgc ccacctaccg cttccagcgc gagcactact ggcccgagac 85680

gggcatcccc ctgcccggcg acaccgctgg gctcgggctc gccgccgcgg gtcatccgct 85740

gctgggtgcg gccgtgacac tcgcggacgc cgacggatgc gtcctcaccg gtcggctctc 85800

cctgcggacg catccctggc tcgcggacca cgccgtcatg gggtccgtac tgctcccggg 85860

aacggctctc gtcgaactgg ccctgcatgc gggcgagcgc gtcggaaccc gtgccctgga 85920

cgagctgacg cttcaggccc cgctgatcct gccgaacgag ggcgcggttc agctgcaagt 85980

cgtggtcggt gcgcccgatg ccgcgggcca ccgcacggtg gccgtgtact cccgcccgga 86040

cgccgacggc gaagcgtggg tccggcacgc cgacggactg ctggtggacg aggtccgggg 86100

cgccgccgcc gacctcggcg tctggccccc ggccggtgcg accgccgttc cggtggacga 86160

cgcctacgcg atcttggaga cctcgggggct cgcgtacggc cccctgttcc aggggctgcg 86220

ggcggcctgg cggcgagcag gagagctgtt cgcggaactg gccctgccca cggaggcgca 86280

ggcggacgcc gccgcgttcg ggctgcaccc tgcgctgctg gactcggcgc tgcacaccct 86340

ggcgctgggt gatctgctgt ccggcgcgga cgcggaggaa acgcccggcg ccgcacggct 86400

gccgttcgcc tggcgtggtg tccgcctcca cgcggccggt gccccggcgg tacgggtccg 86460

gctggccgag gccggtcagg gcgcggtgtc gctggaactg gccgactccg cgggtgcccc 86520

cgtcgcctcg gtggattccc tggtactgcg ggcgatgtcg cccgagcagc tcggcgcggc 86580

gagcgccggc cgccaggagt cgttgttcca gatcgactgg gtggagccgg cggccgaccg 86640

gacggcggct gcgaccgatg tcgaacgggc cctggtgggc ccggagctgc ggggtctgga 86700

cgccacgccg tacgccgacc tggccgcgct ggcggccgcg gactccgacg tgcccgaact 86760

cgtgttcatc accacgcgag cggagtcgga gccggagggc ctgccggggа cggtgcacgt 86820

ccgggccgtc gacgcgctca cccacgtacg ggcatggctg gccgaggaac gcttcgcgtc 86880

cgcccggctg gtgttcgtca cccgcggtgc catgaccgtg ggttcggacg aggccgtccg 86940

cgatctcgcg ggtgccgcgg tgtggggtct ggtccgctcc gccggtaccg agcaccccgg 87000

ccggttcgct ctcgtcgatc tcgacgacga cgacgtgctg cccgagcaga ccgtcctgac 87060

ggccctggcc gcaggggaat cggaactggt cgtacgcgag ggatccctcc ttgtgccgcg 87120

cctcgcgcgt gctgctgtcg ttgagggttc cggtcgtgaa ctggacgtcg acggcacggt 87180

gttggtgacg ggtgcgagtg gcaccttggg tggtttgttc gcccgtcatt tggtggttga 87240

gcgtggtgtg cggcgcctgc tgttggtgag tcgtcgtggt ggggctgcgg agggtgctgc 87300

tgaactgggc gccgaactca cggagctggg tgctgatgtg cggtgggcgg cgtgtgatgt 87360

ggccgaccgt gaggcgcttg agtcggtcct ggccgggatt cccgccgagt atccgttgtc 87420

gggtgtggtg cataccgctg gtgtgctgga cgacggtgtg gtgtcgtccc tgaccgctga 87480

gcgcgtgtcg gcggtgctgc gtccgaaggt ggacgcggca tggaacctgc atgagctgac 87540

ccgtggcctg gatctttctc tcttcgtgtt gttctcgtcg gctgccggtg tgttcggtgg 87600

tgccggtcag gcgaactatg cggcggcgaa tgtgttcctg gacgctctgg cccagcaccg 87660

cagggcccag ggtctggccg cgacctccct tgcgtggggt ctgtgggctg agccgggtgg 87720

tatggcgggc gcgctggacg ctgatgatgt gtcgcgtctg ggccgtggtg gtgtcagcgg 87780

gctgtccgcg ggggagggtg tggcgttgtt cgacgcggca tccgcgtccg aacaggcctt 87840

gttcgttccc gtgaagctgg acctggccgc cctgcgcgcc caggcgggta gcgggatgct 87900

gccgccgctg ctcagcggtc ttgtccgtac ccccacccgc cgcgccgcgg gcaccgccaa 87960

cgctgcggta tccgccccgg gggaccgcct cgccggattg tccgccgctg aacaggtggc 88020

gcacgtactg gagttggtcc gtactcaggt tgccgcggtg ctggggtacg cctccccgga 88080

ggcggtcgag aaggacagct cgttccgcga gctgggcttc gactcgctga ccgccgtcga 88140

gctgcgcaac ctgctcggcg cggcgacggg gctgcgcctg cccgccacgc tcgtcttcga 88200

ctacccgacc tcagcggtcc tggccgacca cctgcggtcg gagctggtcg gaacggcgcc 88260

cgtgacatcg gctccggtcg ttctcgcggc ccgggacgat gacgagccca tcgcgatcgt 88320

gggcctcggc tgccgctacc ccggcggcgt ggagagcccg gacgacctct ggcggctcgt 88380

cctggaaggc cgggatgcca tcacggagtt cccggaggac cggggctggg acgtggacgc 88440

gctgttcgac gccgaccccg accagcaggg tacgagttat gcccgcgagg gcggcttcgt 88500

ccgcgacgcg ggccacttcg acccggcgtt cttcgggatc tcgccgcgcg aggccgtggc 88560

catggacccg cagcagcgac tcctcctcga aacctcgtgg gaggcgttcg aacgggcggg 88620

catcgacccg gcggccctgc gcggcagccg gaccggcgtc ttcgcgggtg tgatgtacca 88680

cgactacgct tcccggctca cggccctccc cgagggcgtc gagggcttcc tcggcacggg 88740

caacgcggcg agcgtcatct ccggacggct gtcgtacgcc ttcggcctgg aaggcccggc 88800

catcaccgtc gacacggcct gctcgtcctc gctggtcgcc ctgcacctgg cggtgcaggc 88860

gctccgcaac ggcgagtgtt ccctcgctct cgcgggcggt gtcacggtca tggcgacccc 88920

cgctgccttc gtggagttca gtcgccagcg cgggctcgcg gccgacggcc ggtgcaaggc 88980

gttctcggcc ggcgccgacg gcacgggctg gtccgagggc gcgggcgtcc tgctggtgga 89040

gcggctctcc gacgcgcggc gcaacggtca cccggtgctc gcggtggtcc gtgggtcggc 89100

gatcaaccag gacggtgcga gcaacggtct gacggctccg aacggtccct cgcagcagcg 89160

ggtgatccgc caggcgctgg ccagcgcggg cctgtcggcg gcggatgtgg acgtcgtgga 89220

ggcgcacggc accggcacca ccctcggcga cccgatcgag gcgcaggcgc tcctcgccac 89280

ctatggccag gagcacacgg acgagcagcc gctgctgctc ggctcgatca agtccaactt 89340

cggccacacg caggccgccg ccggtgtcgc gggcatcatc aagatcgtcc aggcgatgcg 89400

tcacggtgtc gtccccaaga cgctgcacgt ggacgagccc accccgcacg tcgactggtc 89460

ggcgggcgcg gtctcgctcc tcaccgagca ggtggcctgg cccgaaaccg gccgtccccg 89520

ccgcgcggcg atctcttcct tcggcttcag cggcaccaac gcgcacgcca tcatcgagca 89580

ggcccccgac cccgctcccg aggacctgcc cgacgcagga cccgacgtac ggcccgagcc 89640

cgcccggact ccgggcagcc tgccgtggct cctctcggcg aagggcgcgg acgccctgcg 89700

cgaccaggcc gcccggctcc gggcgcatgc catcgggcac cccgagctgt ccctcgccga 89760

catcggctac gccctggcca cgagcaggac cgcgctcgac cggcgggccg ccgtggtcgc 89820

cggggaccgc gaggagttcc tcgcgggact cgcggcgctc gccgagggtg ccacggcggc 89880

cggcctgacg gagggatcac cggccggtgg caagctcgcc ttcctgttca ccgggcaggg 89940

cagccagcgc ctggccatgg gcagggagct gtactccgcc catcccgtct tcgcccgggc 90000

cctggacgcc gtgtgcgacg ggctcgccct ggacgtaccg ctgaagcagg tgctgttcgg 90060

gtccgacgcg gacctgctcg accggaccgc gtacacccag cccgccctct tcgccgtcga 90120

agtcgcgctg ttccgcctgg tcgagagctg gggcctgaag cccgacttcc tggccgggca 90180

ctccatcggc gagatcaccg cggcccatgt ggccggggtg ctctccctcg acgacgcctg 90240

cacgctggtc gccgcccgcg gccggctcat gcaggcactg cccaccggcg gcgtgatgat 90300

cgccgttgag gcatcggagg acgaggtcct gccgctgctc accgaccggg tgagcatcgc 90360

cgcgatcaac ggcccccagt cggtcgtgat cgcgggtgac gaggccgacg cggtggcgat 90420

cgcggagtcc ttcaccggtc gcaagtccaa gcggctcacg gtcagccacg ccttccactc 90480

gccgcacatg gacggcatgc tcgacgcctt ccgcgaggtc gccgagggac tgtcgtacgg 90540

gaccccgctc atccccggtcg tctcccacct caccgggacc ctggtcaccg acgagatgcg 90600

gtcgccggac ttctgggtcc ggcacgtccg cgaggcggtc cgcttcctgg acggcatccg 90660

cacgctggag gacgcgggcg tcaccacgta catcgaactc ggccccggcg gcgtcctctc 90720

cgcgatgggt cagtcgtgcg tcacgcgcga cgacgcggcc ttcctcccgg ccctgcgcgc 90780

ggaccgctcc gaagaggaga cgctcacctc ggccgtcgcc cgggcacacc tgcgcgggat 90840

caccgtcgac tgggacgcgt actactccgg caccggcgcc cggcgcgtcg acctgccgac 90900

gtacgccttc cagaggcagc gctactggct ggaggccccc gcccacgccc ccggcgggga 90960

cgtgacgtcc gccgggctcg gctccgcggg gcacccgctc ctcggcgcgg ccgtcgaact 91020

gccggactcg gacgggttcc tgttcaccgg gcggctctcc ctgcgcaccc accccctggct 91080

cggcgaccac agggtggcgg gcaccgtcct gctgccgggc gccgcgctgc tggaactcgc 91140

cgtgcgcgcc ggggaccacg cgggctgcga tctgctggag gacctcacgc tggaggctcc 91200

gctcgtactg cccgaggcgg gcggggtaca gctgcggctc gtcgtggccg aacccgacgc 91260

gtcgcgcagg cgggtgttcc acatctactc ccgcccggag gacgcggcct tcgaggagcc 91320

gtggacccgg cacgccggcg gtgtcctggc cgtcgagggc gcgcacccgg ccgaggcgga 91380

gtccgagtgg ccgcccgccg gagccgtccc ctgcccggtg gaggacctct acccgtcgct 91440

cgacgccatc gggctcggat acggtcccgc gttccgcaat ctgctgctgg cctggaagcg 91500

cggcgacgag gtgttcgccg aggtcgctct cggcgaggac cggcggaccg aaggcgccct 91560

ctacgggctc cacccggcgc tgctcgacgc cgccctgcac gcggtcggcc tcggggactt 91620

cttccccgac gggcccgagg gcgcgcggct gccgttctcg tgggacggcg tgcggctgca 91680

cgccgtgggc gccgcggcgc tccgggtacg gatggcaccg gccgggcagg acgcggtcac 91740

gctggccgtc tccgacgaaa cgggccggcc ggtcctcacc gtcgactcgc tcgtcctgcg 91800

tccgctggcc ctcgatggtc cgggcgggct cggcggagcg ggccggggac cgggttcggt 91860

gcgcgacgcg ctgttccagg tcgactggca cgcgctgccg ctgcccgagg cgcagtcacc 91920

ggccgaaggc cgctgggccc tgctcggcgg cgacccgctg aagctggccg ccgcgctgga 91980

gcgcaccggg gtcctggagc cgggcgcgct gttcggcacg gcctccgagg acaccggcgg 92040

gcacctcgc gacctgtccg ccctggcgga cgcggtcgag ctggccgagg cactcgggga 92100

gcccgcgccc gagaccgtcc tcgtctccct ggcacccgac ctcgccgcca cgggcggcct 92160

cgcgtcggcc gcccaccgcg ccgccgcgga cgcgctggag ctgatccagg cctggctggc 92220

ggacgagcgg ctcgccggtt cacggctggc cctcgtcacg cggggcgccg tcgccacgga 92280

ccccgacgcg gacgtggacg acctcgcgca cgccgcggtg tggggactgg tgcgctccgc 92340

gcaggccgag caccccggcc ggctggttct ggtcgacctc gacgacgagg acgactccta 92400

ccgggccctg cccgccgcgc tcgacaccga tgagacccag ctcgccgtgc gcgacggggc 92460

cgtcctggcc ccgcgtctgg cgcgagcggt catcgccccg gcaacggatg cggcggcccc 92520

ggacgttgcc ccggacccgg agggcaccgt cctcatcacg ggcgccagcg gcaccctcgg 92580

cggcctgctg gcccggcacc tggtgacgga gcacggtgtg cggcatctgc tgctcaccag 92640

ccgcaggggc gccgctgccg aaggcgccac ccaactcgca gacgaactcg tcacgttggg 92700

tgcgcaggtc acctgggcgg cgtgtgacgc ggccgaccgg gacgcgctgg ccgcgctgct 92760

ggagtccgta cccgcggccc atccgctgac ggccgtcgtg cacaccgccg gtgtgctgga 92820

cgacggcacg gtcgagtcgc tgaccgccgg acggatggcg acggtgctgc ggcccaaggt 92880

cgacgccgcg tggaacctgc acgaactgac ccacggactc gacctggccg cattcgtcct 92940

gttctcctcg gcggccggtg tgttcggcaa cgccgggcag gccaactacg cggcgggcaa 93000

caccttcctg gacgccctcg cccagcaccg ccgcgcccag ggcctcacgg ccgtctcact 93060

ggcctggggt ctgtgggacg acgaggcggg catggcagcc accctcgacg agcaggaccg 93120

gcggcgcctg agccggggca gcatgaaccc gctgtcggtg gccgaggggc tcgcgctctt 93180

cgacgccgcg ctgccgggcg gggcatcctc cggcgccgtg cccgaggggcg cgcggaccgc 93240

gagcgtactc gtgcccgcgc ggctcgactt ggccgtgctc caggcccaag tgggggatct 93300

cgtaccgccc ttgctgcgcg gcctgctccg tactccggta cggcgcaggg cgagcggcgc 93360

ggcggccgac gcgcccgact cgctggcgca gcggctcgcc caactgccgc ccgccgaacg 93420

ggaccgggtg ctgctcgacc tcgtctgcac ccaggtggcc caggtgctgg gccacagcgg 93480

cgcggccgcc atcgaaccgg gaagcgcctt caaggaactc ggcttcgact cgctgaccgc 93540

ggtggagctg cgcaaccggc tcggtgccgt gacggggctg cgcctccccg ccacgctcat 93600

cttcgactac ccgacccccg aagcgctgag cggacatctg cgctccgcgc tgcccctcga 93660

cgaggacgga ccgtccgtct tcagcgaact cgaccggctg gagagcgcct tgggcgcggc 93720

ggacgcggac agcgtcacgc gttcacggat cacgatgcgc ctccaggccc tgatgaccaa 93780

gtggaacgac gcacaggacg cgaacggcgg cgccccccgac gaggacgccg acgacggcgc 93840

cctcgaaacg gcgaccgacg acgagctgtt cgacctgctc gacaacgagc tcggcgcctc 93900

ctgagaaacc gcgcggcgcg cctcccttcc gggccttccg ggcgggggcc gcgccgcccc 93960

gcaccaccgc aacagccacg ggatcccgca cgccgggacc ccgggccacc cagacgaccg 94020

accgtacaac cgcctctctg gcatggagcc cacgcaatgg tgaacgagga caagcttcgc 94080

gactacctca agcgggcgac cgccgatctg cgccaggccc gcaggcggct gcgcgaggtc 94140

gaggacaaga accaggaacc catcgccatc gtcgcgatga gctgccgcta ccccggcggc 94200

gtccgcagcc ccgaggacct gtggcggctc gtggagaacg gcgacgacgc cgtctccggc 94260

ttccccgtcg accgcggctg ggacgtggag gcgctctacg acgccgaccc cgacagctcc 94320

ggatccagct acgtcagcga gggcggcttc ctctacgacg ccgcgagctt cgaccccgcc 94380

cccttcggga tctcgccgcg cgaggccctc gccatggacc cgcagcagcg gctgctcctc 94440

gaagcgtcct gggaggcgtt cgagcgcgcg ggcatcgacc cgtcgtccgt gcgcggcagc 94500

cggacggccg tgttcgccgg tgtgatgtac cacgactaca ccgcgcgcct cgattccgtg 94560

cccgagggcg tcgaaggatt cctcggcacc ggcagctcag gcagcatcgc ctcgggccgg 94620

gtggcctaca cgttcggcct ggagggcccg gcggtcaccg tcgacacggc ctgctcgtcc 94680

tcgctcgtca ccctgcacct ggccgtccag gcgctgcggg ccggcgaatg ctcgatggcg 94740

ctcgcgggcg gtgtcaccgt catggcgacc cccgcgacct tcaccgagtt cagccgccag 94800

cgcggcctcg cgccggacgg gcgctgcaag cccttcgcgg ccgccgcgga cggtacgggc 94860

tggggcgaag gcgtcggcat gctcctcgtc gagcgccttt cggacgctca gcgcaacgga 94920

catccgatcc tcgcggtggt ccgcgggtcg gcgatcaacc aggacggtgc gagcaacggc 94980

ctgacggctc cgaacggtcc gtcgcagcag cgcgtcatcc accaggcgct caccaacgca 95040

cggctgtcgg ccgcggatgt ggacgtcgtc gaggcgcacg gtacggggac gaccctcggc 95100

gacccgatcg aggcgcaggc cctgctcgcc acctacggcc aggaccgccc ggccggacgc 95160

ccgctgctgc tcggctccat caagtccaac atcggccaca cccaggccgc cgcgggtgtc 95220

gcgagcatca tcaagatggt cgaggcgatg cgtcacggag tggtccccaa gaccctccac 95280

ctcgacgagc cgactccgca cgtggactgg gaggcgggcg ccgttctccct gatcggcgag 95340

aagatcgcct ggccggagac cggtgaactc cgtcgtgcgg gtgtgtcgtc gttcgggttc 95400

agcgggacga acgcgcatgt gatcgtcgag caggctccgg tggtcgagga ggtggcgggg 95460

gatccggccg gtgaggtcga gggttcggaa ctcgcggtgg tgccgtgggt gttgtcgggc 95520

aagagtgcgg gggcgttgcg ggcgcaggcg gagcggttgt cggggtggct cgccggtgct 95580

tcggctgcgg gtgtggcgtc ggttgacgtg ggctggtcgt tggcgtcgtc gcgggccggg 95640

ctggaacacc gggctgtggt gctgggcgat cacgcggccg gtgtgggggc ggtggcgtcg 95700

ggtgtgatgg ccgcgggtgt ggtgacgggg tcggttgtcg gcgggaagac cgcgttcgtg 95760

ttcccggggc agggctcgca gtgggtgggt atggcggtgg ggttgctgga ttcctcgccg 95820

gtgttcgctg cgcgggtgga tgagtgtgcg aaggcgttgg agccgttcac tgactggtcg 95880

ttggtggatg tgctgcgggg tgtggagggt gcgccgtcgt tggagcgggt ggatgtggtc 95940

cagcctgctc tgttcgcggt gatggtgtcg ttggcggagg tgtggcgggc tgctggtgtg 96000

cgtcctggtg cggtgatcgg tcattcgcag ggtgagatcg ctgcggcgtg tgtggcgggg 96060

atcttgtcgc ttgaggacgc cgcgcgagtg gttgcgttgc gcagtcaggc gatcggccgg 96120

gtcctggcag gtctcggcgg gatggtgtcg gtgccgctgc ccgcgaaggc agtacgagag 96180

ctgatcgctc cgtggggtga gggccggatc tcggtggccg cggtgaacgg gccgtcctcg 96240

gtggtcgttt cgggtgaggc cgccgccctg gacgagatgc tggcctcgtg cgagtcggag 96300

ggtgtgcggg cgaagcggat cgcggtggat tacgcgtcgc attcggctca ggtggagttg 96360

ctgcgggaag agcttgctga gctgctggct ccgattgttc cgcgcgctgc tgaggtgccg 96420

ttcttgtcga cggtgacggg tgagtgggtg cgaggcccgg agctggatgc tggttactgg 96480

ttccagaatc tgcgccggac ggtggagttg gaagaggcga cgcggacgtt gctggagcag 96540

ggcttcggtg tgttcgtcga gtcgagcccg cacccggtgt tgagcgtggg catgcaggag 96600

acggtcgagg acgcgggccg ggaggcggct gttctgggtt cgctgcgtcg tggtgagggg 96660

ggtctggagc gtttctggct gtcgctgggt gaggcctggg tccgtggcgt ggctgtcgac 96720

tggcatgccg tgttcgcggg tacgggtgcc cggcgggtgg acctgcccac ctacgccttc 96780

cagcaggagc actactggct cgaaagcggc accgccgagg acgtcacggc caccgcccac 96840

cccgtcgacg ccgtcgaagc ccgcttctgg gaggccgtcg agcgccagga cgtggcggcg 96900

ctcaccgccg agctggacgt ggacgagaac gagaacctca ccgcgctgct gcccgcgctg 96960

tcgtcgtggc gtcggcagag ccgtgagcgg tccgccgtgg acggctggcg ctaccgggtg 97020

acctggaagc ccgcgccgga gcccacgacg gcccgcctct ccggcacctg gcttgttgcc 97080

gtcgccgagg gcgcgccggg tgatgagtgg acgtccgctg tcctgcgtac gctcgccgaa 97140

cacggcgccg acgtacggca gatcacggtc gcccggaccg aggacacccg ggccggtctc 97200

gccgagcgga tacgtgacgt actcgcggac ggtcccgcgg tgtcgggagt cttgtccctg 97260

ctgaccccgg cgggggccga cgagccgttc caggtctccg cgcccggcgg tgtgatcacc 97320

accctgtccc tcgtccaggc gctcggcgac gccgaggtgg ccgcaccccct gtggtgcgtc 97380

acgcgcggcg ccgtcgccac cggccgttcc gagcaggtgg ccgaccccgc gcaggctccg 97440

gtctggggcc tgggccgggt gaccgcgctg gagcacggcg agcgctgggg agggctgatc 97500

gacctgcccg gcacggacgc cgtggacgac cgggcactcg cccggctcgc gggcgtcctc 97560

gccggtgacg ccgccgagga ccaggtggcg gtgcgcgcct ccggcctctt cgtacgacgg 97620

ctcgtacgcg tccgtctcgc cgagacgccc gtcgtacggg agtggcgtcc gcagggcacc 97680

accctggtca cgggcggtac gggcgcgctg ggcgcgcacg tggcccgctg gctcgctgag 97740

aacggcgccg agcacctgct gctcaccagc cgccggggcc ccgacgcgcc cggagccgcc 97800

gcactccgcg acgaactcac cgccctcggc gcccaggtca ccatcgcggc ctgcgatgtg 97860

agcgaccggg acgccgtcgc ggccctcatc gccgcggttc ccgccgacca gcccctcacc 97920

gccgtcgtgc acacggcggc cgtcctcgat gacggggtca tcgaggcgct cacgcccgag 97980

cagatcgagc gcgtcctgcg ggtgaaggtc gacgcgacgc tgcacctgca cgaactgacc 98040

cgcgagctcg acctgtcggc gttcgtgttc ttctcgtcct tcgccgccac cttcggcgcc 98100

cccggccagg gcaactacgc gccgggcaac gcgttcctgg acgccttcgc cgagtaccgc 98160

cgggcatccg gactgcccgc cacctccatc gcctggggcc cttggggcga cgggggcatg 98220

gccgagggcg cggtcggtga ccggatgcgc cgccacgggg tcatcgagat gtcgcccgag 98280

cgtgccgtcg ccgcactcca gcacgccctg gaccgcgacg agacgaccct gaccgtcgcc 98340

gacatggagt ggaagcgctt cgtcctcgcc ttcacctccg gccgcgccag gccgctgctg 98400

cacgacctgc ccgaggcgcg ggaggtcatg gacgccacgc gcacggaggc ggcggaggac 98460

accggcagcg ccgccgcgct ggcccagcag ctgaccggcc ggccccgaggc cgaacaggag 98520

cgactgctcc tcgaactggt ccgcaccgcc gtcgccgccg tcctcggcta cgcgggcccc 98580

gacgcggtcg aggcgggccg ggccttcaag gagctgggct tcgactccct cacctccgtc 98640

gaactgcgca accgcctgaa cgcggccagc ggcctcaagc tgccgcccac cctcgtcttc 98700

gaccacccga cgcccaccgt cctcgcccgg cacctgcggg ccgagttctt cggccagggc 98760

gccgcggccg ccgtgcccgt gccgatggcc gcggtctccg acgacgagcc gatcgccatc 98820

gtcgcgatga gctgccgctt ccccggcggg gtccgcaacc ccgaggagct gtggcagctg 98880

ctcacctccg agggtgacgg gctgtcccag ttccccctgg accgcggctg ggacgtcgac 98940

gcgctgtacg accccaaccc cgacgcgcaa ggcacctcgt acacgcggga gggcggcttc 99000

ctgtccgacg ccgcggcctt cgactcctcg ttcttcggga tctcgccgcg cgaggccctc 99060

gccatggacc cgcagcagcg gctgctcctc gaaacctcgt gggaggcgtt cgagcgggcg 99120

ggcatcgacc cgcagaccct gcgcggcagc cagtccggtg tgttcgtcgg caccaacggc 99180

tctgactact ccaacctcgt acgggcgggg gcggacggcc tggaggggca cctggccacc 99240

ggcaacgcgg gcagtgtcgt ctccggccgg ctctcctaca ccttcggtct cgaaggcccg 99300

gccgtcaccg tcgacaccgc ctgctcggcc tccctcgtcg ccctccacct cgccgtgcag 99360

gccctgcgca gcggtgaatg ctcgctcgcc ctggccggtg gcgtgacggt gatgtccacg 99420

ccgggcacct tcatcgagtt cagccgtcag cgcggactct ccaccgacgg ccgctgcaag 99480

gcgttctcct cggacgccga cggattcagc cccgcggagg gcgtcggcgt gctcctcgtc 99540

gagcgccttt cggacgctcg gcgcaacggg catccgatcc tcgcggtggt ccgtgggtcg 99600

gcgatcaacc aggacggtgc gagcaacggt ctgacggctc cgaacggtcc gtcgcagcag 99660

cgcgtcatcc ggcaggccct cgccaacgca cggctgtcgg ccgcggatgt ggacgtcgtc 99720

gaggcgcacg gtacgggtac gacgctgggt gacccgatcg aggcgcaggc cctgctcgcc 99780

acctacggcc aggaccgccc ggccggccgg ccgctgctgc tcggctccat caagtccaac 99840

atcggccacg cccaggcggc ggccggtgtc gcgggcgtca tgaagatggt gctcgccatg 99900

cagcacggag tgctgccgca gagcctgcac atcgccgagc ccacgccgca cgtcgactgg 99960

agcgcgggcg aggtcgccct gctcaccgag gagcgggcct ggcccgagac cggccgcccc 100020

tggcgggcgg gcgtctcgtc gttcggcttc agcggcacca acgcccacgc catcatcgag 100080

caggctccgg ccgaagcggg atccgacgac gaccgggaga cccctgagcc gtcggcccaa 100140

cccctactgg tcgcgcccac ccgggacgac tccgcgtccg cccgggacga ctccgcgtcc 100200

gccccggacg gctccgtatc cggcccggac gactccgtgt ccgaccgtcc cggcgtgctg 100260

ccctggaccc tgacggccaa gaccgagaag gcgctgcaag gccaggccga acgcctgctg 100320

acccagctca ccacccgctc tgacctgcga cttgtcgatg tcggccactc cctggcgacg 100380

acccgtaccg cgctcgacca gcgcgccgtc ctcatcggac gggaccgccc cgactacctc 100440

ggagccctga ccgcactcgc ggcgggggac acctcccccc tgctggtgca gggggcggtc 100500

gtcgggggga agacggcgtt cgtgttcccc ggacaggggt cgcaatgggt aggcatggcg 100560

gtggcgctgt tggacgcttc acccgtgttc gctgcccgag tggatgagtg tgcgaaggcc 100620

cttgagccct tcaccgactg gtcgctgcgc gatgtactgc gcggcgtcac aggcgcgccg 100680

tcgttggacc gcgtggatgt ggtccagcct gctctgtttg cggtgatggt gtcgttggcg 100740

gaggtgtggc gggccgctgg tgtgcgtcct gatgcggtga tcggtcactc gcagggcgag 100800

atcgctgccg cgtgtgtggc gggcatcttg tcgcttgagg acgcggcgcg agtggtcgcg 100860

ttgcgcagtc aggcgatcgg ccgggtcctg gcgggcctgg gcgggatggt gtccgtggca 100920

ctgccggcga aggctgtgcg ggagctgatc gctccgtggg gcgaggaccg gatctcggtg 100980

gccgcggtga acgggccttc ctccgtggtc gtttccggtg agaccgccgc cctggacgag 101040

ctgctggcct cgtgcgagtc ggacggcgtc cgggcgaagc ggatcgcggt ggattacgcg 101100

tcgcattcgg ctcaggtgga gttgctgcgt gaggagcttg ctgagctgct ggctccgatt 101160

gttccgcggg ctgccgaggt gccgttcctg tcgacggtga cgggtgagtg ggtgcgcggt 101220
ccggagctgg atggcgggta ctggttccag aacctgcgtc ggacggtgga gttggaagag 101280
gcgacgcgga cgttgctgga gcagggcttc ggtgtgttcg tcgagtcgag cccgcacccc 101340
gttctgacga tgggtgtgca ggagaccgtc gaggacgcgg gccgtgacgc ggctgttctg 101400
ggctcgctgc gtcgtggtga ggggggtctg gagcgtttct ggctgtcgct gggtgaggcc 101460
tgggtccgtg gcgtgggtgt ggactggagt gccgtgttcg cgggcacggg tgcccggcgg 101520
gtggatctgc ccacttacgc cttccagtcg cagcggttct ggccggaggc cgcgcccatc 101580
gaggctgtgg cggtgtcggc ggagagtgcg atcgatgcgc ggttctggga ggccgtcgag 101640
cgcgaggatc tcgaagcgct gaccgctgag ctcgacatcg agggcgacca gccgctgacc 101700
gcgctgctgc ccgcgctgtc gtcgtggcgt cggcagagcc gtgagcactc gacggtggac 101760
ggctggcgct accgggtcac ctggaagccg ctggccgagg ccaagacctc tcgcctctcc 101820
ggtacttggc tggtcgtcgt tcccgagaac ggcccggccg acgagtggac gggggccgtg 101880
ctgcgcgtgc tcgccgaccg cggcgcggag gtccgtactg tgaccgtccc ggccgacggg 101940
gccgatcgtg accggctcgc cgccacgctg aaggccgaga cggacggggc cgctccggcc 102000
ggagtgctgt ccctcctcgc ccttgccgtc gaaagcgctg aactccgtac gcacaccggg 102060
ctcctcgcca ccgccgcgct cgtccaggcg cttggtgacg ccgatgtggc cgcacccctg 102120
tggtgcgtca cgcgtggcgc tgtctccgtc gcccgtacgg agcggctcca ggacccggcg 102180
caggcgctcg tgtcgggctt cggacgcacg gtcgccctgg agtacccgga ccgttggggc 102240
ggtctcgtcg acctgccgga gcaggccgac ggccgtacgc tcgaacgtct tgcgggtgtg 102300
ctggccggtg acggttccga ggaccaggtg gcgctgcgcg cctcgggtct cttcggccgg 102360
cgtctggtcc acgcacccct cgccgacacc gccgcggtac gggagtggcg tccgcagggc 102420
acgaccctgg tcaccggtgg tacgggtgcg ctgggcgcgc acgtggcccg ctggctcgct 102480
gagaacggtg ccgagcactt gctgctcacc agccgccggg gcccggacgc gcccggtgcc 102540
gccgaactcc gcgacgaact cacggccctc ggcgcccagg tcaccatcgc cacctgcgac 102600
atggccgacc gggacgccgt cgcggccctc atcgccgccg ttcccgccga ccagcccctc 102660
accgcggtga tgcacacggc cggtgtcctc gacgacggcg tgatcgacgc gttgactccg 102720
gagcggttcg ggacggtgct cgcccccaag gcggacgcgg ccctcaccct ccatgagctg 102780
acccgcgagc tgggcctctc ggcgttcgtc ctcttctccg gtgtcgcggg cacgctcggc 102840
gacgcgggac agggcaacta cgccgccgca aactcctact tggacgccct cgccgagcag 102900
cgtcacgccg acggcctcgc cgccacctcg gtggcctggg gtcgctgggg cgacagcggg 102960
ctcgccgcgg gcggtgcgat cggtgagcgg ctcgaccgcg gcggggtgcc cgccatggca 103020
ccccgctcgg cgatccgcgc gctgcagctg gccctcgacc acgcggaggc ggccgtcgcc 103080

gtcgccgaca tccagtggga gcggttcgcg cccggctaca cggcggtgcg gcccagcccg 103140

ttcctcggtg acctgccgga ggtgcggcag ctcgccgcgt ccgctccggc ggccggtgaa 103200

gcgggcgggg actccccggc cgaggcgctg cgccgacggc tcgccgtcat gccgcaggcc 103260

gaacaggccc tggccgtcct cgaactggtc cgctcccacg cggccaccgc gctgggccac 103320

cccacgaccg acgaggtggg cgcgggccgc gcgttcaagg agctcggatt cgactccctg 103380

atcgcgctgg aactgcgcaa ccggctcaac gcagccaccg ggctgaggct cccggccacg 103440

ctcgtattcg accacccgac cccgacgatc ctggccgagt tcctccgggc cgagatcacc 103500

caggacggca gtgccggggc cgccccgggc atcacggaac tcgaaaagct ggagtccgcg 103560

ctgtccgttc tcgacccgga cagtgaaacg cgtaccgata tcgcactgcg cctgcaggca 103620

cttctcgcga aatggggtga accgcacatc gaatcaagtg gcgaggccgt gaccgagaaa 103680

ctccaggagg ccacgcccga cgaactcttc gaattcatcg agaaagagtt cggtatttag 103740

cacagcggac agcaggcagt agcagcgcaa gggtttgtga cgagaagcat gggtgaggtt 103800

ccaatggcag atcaggacaa gatcctcggt tacctgaagc gggtgacggc cgatctgcac 103860

cagacgcgcc agcgccttcg tgaggtcgag gcccaggagc cggagccgat cgcgatcgtc 103920

ggcatgagct gcaggttccc cggcggcatc gagtcgccgg agggcctgtg ggacctggtg 103980

gccggtgggc gggacgcgat caccgatttc cccaccgacc gtggctggga catcgagtcg 104040

ctgtacgacg ccgaccccga ccagcagggc acctcgtaca cccgtgaggg cggattcctc 104100

gacggcgtcg ggaagttcga cgcgtccttc ttcgggatca gcccgcgcga aaccctcggc 104160

atggacccgc agcagcgcct gctcctcgaa acgtcctggg aagccttcga aagagccgga 104220

atcgacgcgg ctaccctgcg cggcagcaag gccggtgtct tcataggcac caacggccag 104280

gactatccgg agctgctgcg cgaagtcccc aagggtgtcg agggatatct cctcaccgga 104340

aacgcggcca gcgtcgtctc cggccgcatt tcctacacct tcggcctcga aggcccggcc 104400

gtcaccgtcg acaccgcctg ctcggcctcg ctcgtcgccc tgcacctcgc cgtccaggcg 104460

ctgcgcaacg acgagtgctc gctggcgctg gcgggcggtg tcaccgtgat gtcgagcccg 104520

cgcgcgttcg tacagttcag ccgccagcgc gggctcgcgc ccgacggacg ctgcaagccg 104580

ttcgccgacg gggccgacgg caccggctgg ggcgagggcg tcggcatgct gctcgtcgag 104640

cggctctccg acgcccgcag gaacggtcat cccgtcctcg ccctcgtgcg cggctcggcg 104700

atcaaccagg acggcgcgag caacggcctg accgcgccca acggcccgtc ccagcagcgg 104760

gtgatccggc aggcgctcac gaacgccggg ctcacccccg cgcaggtcga cgtcgtcgag 104820

gcgcacggca ccggtacgac cctcggcgac ccgatcgagg cgcaggccct gctcgccacg 104880

tacggccaga accgccccga ggggcgcccg ctgtggctgg gttccgtcaa gtcgaacatc 104940

gggcacacgc aggccgccgc cggtgtcgcg ggcatcatca agatggtcct cgccatgcag 105000

cacggcgtgc tgcccgagtc gctccacatc gaccagccgt ccggcaacgt cgactgggcc 105060

gccggtgacg tcaagctgct caccgaggcc gtgccgtggc cgcagaccgg ccagccgcgc 105120

cgcgccggcg tctcctcctt cggcgtcagc ggcaccaacg cgcacaccgt catcgagcag 105180

gccccgcccg ccgacgacgc gccggagacc ggcgcggaca ccgcacccac cgccgaggcg 105240

ccggaggcgg cctccgcgga cgcttccgag gccgggacgc cgaccggtgc caccggcccg 105300

gtgccggtgc tcgtctcggg ccagagcgac gccgcactgc gcgcccaggc cgagcgcctc 105360

gccgcccacc tgcgcgccca ccccggactc ggggccgaca ccggaaccct gaccgacctc 105420

ggtttctcgc tcgccaccag ccgctcctcg ctcgaccgca gggccgtcct gttcggcgac 105480

cgggacagcc tgctcgccga cctcagcgcc ctcgccgagg gcgagcagcc cgccggcccg 105540

gtcctcggcg cggtgggcga gggcaagacc gccttcctct tcaccggcca gggcagccag 105600

cgcctgggca tgggacgcga gctgtacgcc acgcatcccg gcttcgcccg cgccctcgac 105660

gaggtccgcg cggaactgga ccagcacctc gaacgccccc tgttcgacgt cctgttcgcc 105720

gccgaaggca cccccgaggc ggacctgctc gacgagaccg cctacaccca gagcgccctg 105780

ttcgccgtcg aggtcgccct gttccggcag ctcgaacagt ggggcgtcgg cgccgacttc 105840

ctcatcggcc actccatcgg cgaactcgcc gccgcccacg tctccggcgt gttcacccta 105900

gccgacgcgg ccaagctcgt cgccgcccgc ggccgcctca tgcaggcgct gcccgccgac 105960

ggcgcgatga tcgccgtcga ggccaccgag gacgaggtcg caccgctgct caccggccgg 106020

gtgagcatcg ccgccgtcaa cggcccccgc tccgtggtcg tctcgggcga cgaggacgcc 106080

gccacggcgc tcgccgagac cctgcgcgca cggggccgca ggacgaagcg gctcacggtc 106140

agccacgcct tccactcgcc gctgatggac ggcatgctcg acgcgttccg tgaggtcgcc 106200

gagagcgtcg cctacgcgcc gcccgtcatc ccgatcgtct ccaacctgac cggcgcctcc 106260

gtcaccgcgg aggagatctg cgccgccgac tactgggtgc gccacgtccg cgaggccgtc 106320

cgcttcctcg acggagtccg caagctctcc gcgcagggcg tcaccacctt cgtcgaggtg 106380

ggaccgggcg gggtcctcac cgccctggcg caggagtgcg tcaccggcca ggacgccgtc 106440

ttcgtgcccg tcctgcgcgg tgaccgcccc gaggcggccg ccttcgcgac ggccgtcgcc 106500

caggcccatg tccacggtgt ggccgtcgac tggtccgccg tcttccgccgg gcgcggagcc 106560

acccgcatcg acctgccgac gtacgccttc cagcgcgagc tgtactggcc cgagcagccc 106620

accgcctggg cgggcgacgt caccgccgcc gggatcggcg ccgccgacca cccgctgctg 106680

ggcgcggcca tcgccctggc cgacggcgac gggcacctgt tcaccgggcg gctctcgctg 106740

gccacccacc cctggctcgc cgaccacacg gtgatggaca ccgtgctgct gcccggcacc 106800

gccttcgtcg aactcgccct ccaggcgggc gaccacaccg gctgcgacct gctggacgaa 106860

ctcaccctgg aagcaccgct ggtgctgccc ccgcacggcg gggtgcagat ccagctcgcc 106920

gtgggcgcgc ccgacgccga gggccgccgc tcgctgacac tgcactcccg gcccgaggac 106980

gccgccgacg acacctgggg agagggcgcc tggacgcgcc acgccaccgg cttcctcgcc 107040

accgccgccc agggcgcccg cgagcccctc gccgacctca ccagctggcc gccgaagaac 107100

gccacgaagg tcgacgtaga aggcctgtac gcgtacctca ccgagtccgg cttcgcctac 107160

ggtccggtct tccagggcct gaccggcgcc tggcagcgcg gcgacgaggt cttcgccgag 107220

gtccgcctgc cggagcaggc gcacgccgag gccgccctgt tcggtctgca tcccgcgctg 107280

ctggacgccg cgctgcacgc cgtcggcatc ggctccctcc tggaggacac cgaacacggc 107340

aggctgccgt tctcctggag cggagtctcc ctgcgggcgg tcggcgcccg tgccctgcgc 107400

gtccggctcg cccccgcagg caacgacacc gtgtcggtga ccctcgccga cgagaccgga 107460

gcgcccgtcg ccgccgtcga cgcgctgctg ctgcggcccg tctccccgga ccaggtgcac 107520

gccgcccgca ccgccttcca cgactcgctg ttccgcgtgg agtggaccgg tacgcccctc 107580

ccggccgcca ccaccgtcgc cgcgggccag tgggcgctgc tgggcgagcc ccgtacggag 107640

ttcaccgccg cgctgcccac cgccgccacc cacgccgacc tcgccgccct cggcgcggcg 107700

ctggacgcgg gcggcccggt cccgcgggcc gtcatcgtcc cgttctccgc gtccggcgcc 107760

ccctcggcga ctcccgtcga cgccgcgctg cccaccgccg tcgccgacgc cctgcaccgc 107820

accctggagc tcgcccaggc gtggctcgcc gacgaccggt tcgccggctc ccggctcgtg 107880

ttcgtcaccc gcgacgccgt cgccaccacc gccggatccg atgtcgccga cctggcccac 107940

gccccgctgt ggggtctgct gcgctccgcg cagtccgagc accccgaccg gttcgtcctg 108000

ctggacctgg acggacgcga ggactccctg cgggccctgc ccgccgcgct cgccacggcc 108060

gagccgcagc tcgccctgcg cgcgggcaag gccctcgtgc cccggctcgc ccgggtcgcc 108120

gccgcccccg gccaggaggc gcccgcgctc gaccccgacg gcaccgccct ggtcaccggc 108180

gccaccggca ccctcggcgg cctggtcgcc cgccacctcg tcgccgcgca cggcgtccgc 108240

cacctgctgc tgaccagccg gcgcggcgag gccgccgccg gcgccgccga actcgccgcc 108300

ggactgcggg aactgggcgc cgaggtcacc atcgcggcct gtgacgccgc cgaccgcgac 108360

gcgctcgccg cgctcatcgg gtccgtaccg gccgaacacc cgctcaccgc cgtcgtccac 108420

accgccggag tcctcgacga cggcgtcctc gaagcgctca cccccgagcg catcgacgcc 108480

gtcctgcccg ccaaggtcga cgcggccgtg cacctgcacg agctgacccg cgagctggac 108540

ctcgcggcct tcgtcctgtt ctccgccgcc gccggcaccc tcggcggccc cggacaggcc 108600

aactacgccg ccgccaacac cttcctcgac gcgctcgccc accggcgccg cgccgaagga 108660

ctgcccgcca ccgccctcgc ctggggcctg tgggccgaac gcagcggcat gaccggcgac 108720

ctcgccgacg ccgacctgga gcggatctcc cgcgccggag tcgccgccct gtcgtccgcc 108780

gagggcctgg cgctgctgga caccgcccgc gccgtgggcg accccaccgc cgtccccatg 108840

cacctcgacc tggcgtccct gcgccacgcc gacgcgagca tggtccccgc gctgctgcgc 108900

ggcctggtcc gcgcgcccgc ccgcaggtcc gtcgagtccc cgggcgccgc cccggccggc 108960

ggcctcgccg agcgcctgct gcccctgacc gccgccgagc gcgaccggct gctcctggac 109020

accgtccggg tccaggtcgc cgccgtcctc ggctaccccg gccccgaggc cgtcgacccg 109080

ggccgtgcct tcaaggaact cggcttcgac tcgctgaccg ccgtagagct gcgcaaccgc 109140

ctcggctccg ccaccggcgt acggctgccc gccaccctcg tcttcgacta ccccaccccg 109200

aacgcgctct ccgcgttcct gcggaccgaa ctcctcggcg acgccgcgga ctcggccccg 109260

gtcgcggccg tcaccgcccg tgacgacgag cccatcgcca tcgtcggcat gagctgccgc 109320

taccccggcg gggtcaccac ccccgaggag ctgtggcagc tcgtcgccgg ctccgtcgac 109380

gcgatctcgc ccttccccac ggaccgcggc tggaacctcg acgcgctgta cgacgccgac 109440

cccggccggg ccgggacctc gtacacccgg gagggcggct tcctgcacga cgccgccgac 109500

ttcgacccgg acgtcttcgg catcaacccg cgcgaagccc tcgccatgga cccgcaccag 109560

cggctcctcc tggagacgtc ctgggaggcg ttcgagcagg ccgggatcgc cccctcgtcc 109620

atgcgcggca gccgcaccgg cgtgttcgcc ggcgtcatgt accacgacta cctgacccgg 109680

ctcccggccg tgcccgaggg cctggagggc tacctcggca ccggcaccgc gggcagcgtc 109740

gcctccggcc gcatctcgta caccttcggc ctcgaaggcc ccgccgtcac cgtcgacacg 109800

gcctgctcct cctcgctggt cgccctgcac ctcgcggccc aggccctgcg caacggcgaa 109860

tgcgacatgg ccctcgcggg cggtgtcacc gtcatgtcca ccccggacac cttcatcgac 109920

ttcagccgcc agcgcggcct ctccggcaac ggccgctgca agtccttctc cgccgacgcc 109980

gacggaaccg gctgggccga gggcgcgggc atgatcctcg tcgagcggct ctccgacgcc 110040

cgccgcaacg gccaccaggt cctggcggtc gtccgcggca ccgccgtcaa ccaggacggc 110100

gccagcaacg gcctgaccgc cccgaacggc ccctcccagc agcgcgtcat ccgccaggcc 110160

ctcgccaacg cgggcctgac caccgccgag gtcgacgtcg tcgaggcgca cggcaccggc 110220

accaccctcg gcgaccccat cgaggcgcag gccctcctcg ccacctacgg ccaggaccgc 110280

ccggccgggc agccgctgcg gctcggctcc atcaagtcca acatcggcca cacccaggcc 110340

gcggcgggcg cggcgggcat catcaagatg atcctcgcca tgcgccacgg cgtcatgccg 110400

ccgtcgctgc acatcggcga gccgtccccg cacatcgact ggaccgcggg cgcggtctcg 110460

ctgctcaccg aggccgccga gtggccccgac gcgggccgcc cccgccgcgc gggcatctcc 110520

tccttcggcg tcagcggcac caacgcccac gtcatcatcg agcagccgcc cgtcgaggaa 110580

cccgccaccg cgaccgagac cggctccggc accggcctgc ccgccggcac gcccctgccg 110640

ttcgccctct ccggccggac ccccgccgcg ctgcgcgccc aggccgcccg gctgatcggc 110700

cacctcgcgc cgcggcccga ggccgccccc gccgatgtgg cgctctcgct ggccaccacc 110760

cgtaccgccc tggaccgcag ggccgccgtc atcgcgcacg accgcaccga gctcctcgcc 110820

gggctcaccg ccctggccga gggccacgac agcgcccggc tggtccagca caccgccgcc 110880

gacggccgca ccgcgatcct gttcaccgga cagggcagcc agcgccccgg catgggacgc 110940

gagctgtacg agacgtaccc cgccttcgcc gaggcgctgg acgcggtctg cgccgagctg 111000

gacccgcacc tcgaacagcc cctcaaggag gtcctgttca ccgccgacgg cgacctgctg 111060

aaccggaccg gccgcaccca gcccgccctg ttcgcgctgg agaccgccct gtaccggctc 111120

gtcgaatcgt ggggcgtgcg ccccgacttc gtcgccgggc actccatcgg cgagatcacc 111180

gccgcgcacg tcgcgggcgt cctctccctg cccgacgcgg ccacccctggt cgccgcccgc 111240

ggccgcctca tgcaggaact gcccgagggc ggcgcgatga tcgcgctcac cgccaccgag 111300

gacgaggtcc tgccgctgct ggccggccac gaggaccgca tcggcatcgc cgccgtcaac 111360

tcagcctcct ccgtggtcat ttccggcgag gagggcctcg cgctggagat cgccgccgag 111420

ttcgagcggc gcggtcggcg caccaagcgg ctcaccgtca gccacgcctt ccactcgccg 111480

ctgatggacg gcatgctcga cgccttccgc gaggtcgccg agtccctgac ctaccgggcg 111540

cccgccatcc cggtcgtcac gctcctcacg ggaacggtcg ccggggacga actgcgcacc 111600

gccgagcact gggtctccca cgtccgcgag gcggtccgct tcctcgacgg catccgcacc 111660

ctggacgccg agcacgtcac cacctacctc gaactcggcc cgcagggcgt gctgtccggc 111720

ctcggccgcg actgcctcac cgaccccgcc gaccccggccg acaccgccgt cttcgtaccg 111780

gcgctgcgcc gcgaccgcgg cgaggccgaa gccctgaccg ccgcgatcgc cgcggcccac 111840

acccgcggtg tgccgctcga ctggtccgcg tacttcgcgg gcaccggcgc ccgccgcgtc 111900

gaactgccca cctacgcctt ccagcgcgag cggttctggc tcgaagcccc ggccggctac 111960

atcggcgacg tcgaatcggc gggcatgggc gcggcccacc acccgctgct cggcgccgcc 112020

gtcgccctcg ccgacggcga aggattcctg ttcaccggcc ggctctcgct cgacacccac 112080

ccctggctcg ccgaccacgc cgtcatgggc aacgtcctgc tgccgggcac cgccttcgtc 112140

gaactcgcca tccgcgcggg cgaccaggcc ggctgcgacc tcctcgaaga actcaccctc 112200

gaagcaccgc tgatcctcgc cccgcaggcc gcggcacgcc tccagatcgt ggtcggagcc 112260

cccgacgggt ccggccgccg caccctggac gtgtactcca gcgacccgga cgcccccgcc 112320

gacgagccgt ggacccgcca cgccggcggc atcctcgcca ccggggcaca ggcacccgcc 112380

ttcgacctga ccgcgtggcc cccgccgggc gccgaagccg tcggcgtcga cggcctctac 112440

gaacacctcg gccggggcgg cttcgcctac ggtccgctct tccaggggct gcgcgccgcc 112500

tggctcctcg gcgacgacgt gtacgccgag gtcgccctgc ccgacgaccg gcaggccgag 112560

gccgcccggt tcggcctgca cccggcgctc ctcgacgcgg ccctgcacgc caccttcgtc 112620

cagccgtccc ccgacgggga ccagcagggc cggctgccgt tctcctggcg cgatgtgtcc 112680

ctgcacgccg tcggtgcgtc cgcgctgcgc gtccgcctca cccccgacgg ccgggacacc 112740

ctctccctcc agctcgctga caccaccggc gctcccgtcg ccgccgtcgg ccacctgacg 112800

ctgcggcccg tctccgccga ccagctcggc agcgcacgct ccgcacacca cgagtccctg 112860

ttccggatcg actgggccac cgtgccgctg ccgtccgacg cccccgccgc cacggacgag 112920

tgggccgtca tagccgcgga cggaggcacg gacggcggta cggacggagg cacggacggc 112980

ggcatccccg ccgccctccc cgggcgcgtg cacaccggcc tggacgccct cggcgcggca 113040

gtcgacgcgg gcgccccggt gcccgcccac gtcctggtgc accacacccc cgcggccacc 113100

accgccgacg ccgtccacgc ggccacccac gaggcgctcc gcctcgtccg ggcctggctc 113160

gccgacgacc ggttcgccgc gtcccgcctg gtcttcgtca cccgcggcgc gatcgccacg 113220

cagagcgact gggacctcac cgacctgacc cacgcccccg tgtggggact ggtgcgcacc 113280

gcccagtccg agaaccccga ccggttcgtc ctcgccgacc tcgacgccga cccggcctcg 113340

acggacgccc tcgccgcagc cctcgccacc ggcgagccgc agctcgcggt ccgccgtggc 113400

accgtccacg ccccccgcct cgccccgctc cccgccgcca ccccgctgac cccgcccccg 113460

ggcgagtccg cctggcgcat ggacatcgag gacaagggaa cgctcgacca cctcacccтc 113520

gtccccagcc cggagtccgc cgcgcccctg gagcccggcc aggtccgcgt cgccgtccgc 113580

gccgcgggcc tcaacttccg cgatgtgctc aacgccctcg gcatgtaccc cggcgacccg 113640

ggcctcatgg gcagcgaagg cgccggcatc gtcgtggaga cgggccccgg tgtcaccggc 113700

ctcgcacccg gcgaccgcgt catgggcatg ctgcccggct cgttcggccc gctcgcggtc 113760

gtcgaccgcc gcatgatcgc ccccatgccc gagggctgga ccttcgccga ggccgcgtcc 113820

gtacccatcg tcttcatgac ggcgtactac gccctccacg acctcgccgg actgcagggc 113880

ggcgagtccc tcctcgtgca cgccgccgcc ggtggcgtcg gcatggccgc cgtccagctc 113940

gcccgccact gggcgccga cgtctacgcg acggccagcc ccgccaagtg ggacaccctg 114000

cgcggactcg gcctcggcga cgaccggatc gcctcgtccc gcaccctcga cttcgaggag 114060

accttccgca cggccaccgg gggacgcggc gtcgacgtcg tactcgactc gctggcccgg 114120

gagttcgtcg acgcctccct gcggctcctg ccgcgcggcg gacgcttcgt cgaaatgggc 114180

aagaccgacg tccgctcccc gcaggacgtc gccgacgccc acccgggcgt cagctaccag 114240

gcgttcgacc tgaccgaggc cggcctcgac cgcatccagg agatgctcac cgagctgctc 114300

accctcttcc gctccggcgc cctgcgcccc gtaccggtct ccgcatggga cctgcggcag 114360

gcccccgagg cgttccgcta cctcagccag gcacgccacg tcggcaagat cgtgctcacc 114420

ctgccgggcg agtggaactc gcagggcacc gtcctcatca ccggcggcac cggcaccctc 114480

ggcgcggtgg tcgcccggca cgccgtcacc acccgcggcg cccgccgcct gctgctcacc 114540

agtcggcgcg gcgaggccgc cgccggcgcc gccgaactcg ccgccgaact gcgggaactg 114600

ggcgccgagg tcacgatcgc ggcctgcgac gccgccgacc gcgacgcgct cgccgcgctc 114660

atcgaatcca taccgtcaga gcacccgctg acggccgtca tccacaccgc cggagtcctc 114720

gacgacggcg tcgtcgactc gctgaccccc gagcgcctgt ccacggtcct gcgcccgaag 114780

gtggacgccg cctggaacct gcacgagctg acccgtcacc tcgacctggc cgacttcgtc 114840

ctgttctcct ccgccgccgg caccttcggc ggcgccggac aggccaacta cgcggccgcg 114900

aacgtcttcc tggacgccct cgcccgccac cggcacgccc acggcctcgc cgccacctcc 114960

ctggcctggg gcctgtgggc cgaggccagc ggcatgaccg gcgaactcga caccgccgac 115020

aaggaccgga tgacgcgctc cggcgtcctc ggcctctcct ccgaagaggg cgtggcgctg 115080

ctcgacaccg cacggctcac cggcgacgcc ctcctcgtcc ccatgcacct cgacctggcg 115140

ccgctgcgcc ggaccgacgc cagcatggtc cccgccctgc tgcgcggcct ggtccgcgcc 115200

cccgcccgca gggccgtcgg agccaccgcc gccggcgccg gaaccccgct ggtggagcgg 115260

ctcgtacggc tccccgagaa cgagcgcgac ccgctcctgc tcgacctcgt acgccagcag 115320

gtggccgccg tactcggcca cgccaccccc gacgccgtcg aacccacccg cgcgttcaag 115380

gacctcggct tcgactcgct gaccgccgtg gagttccgca accggctcgg cgcgaccgcc 115440

ggcatccggc tgcccgccac gctcgtcttc gactaccccа cccccacggt cctggccggc 115500

tacctcaagg acgaactcct cggctccgag gccgcggccg ccctcccgaa gctcgccgcc 115560

accgccgtcg agggcgacga ccccatcgcc atcgtcgcca tgagctgccg cttccccggt 115620

gacgtccgca ctcccgagga cctgtgggag ctgctcgccg agggccgcga cggcatctcc 115680

gacctcccgg acgaccgcgg ctgggacacc gaggcgctgt acgaccccga ccccgacagc 115740

cccggcacct cctatgccag ggagggcgga ttcttctacg acgcccacca cttcgacccg 115800

gcgttcttcg ggatcaaccc gcgcgaggcc ctcgccatgg acccgcagca gcgcctgctg 115860

ctggagacgt cctgggaggc gttcgagcgg gccgggatcg acccgacggg cctgcgcggc 115920

aagcaggtcg gcgtcttcgt cggccagatg cacaacgact acgtgtcccg gctgaacacc 115980

gtccccgaag gcgtcgaggg ctacctcggc accggcggct ccagcagcat cgcctccggc 116040

cgcgtctcct acaccttcga cttcgaaggc cccgccgtca ccgtcgacac ggcctgctcc 116100

tcgtcgctgg tcgccctgca cctcgcggcc caggccctgc gcaacggcga gtgcacgctg 116160

gccctcgcgg gcggcgtcac catcatcacc acccccgacg tcttcaccga gttcagccgc 116220

cagcgcggcc tcgccagcga cggccgctgc aagccgttcg ccgaggccgc cgacggcacg 116280

gcgtggggag agggcgtcgg catgctgctc gtcgagcggc tctcggacgc ccgccgcaac 116340

ggccaccagg tcctggcggt cgtccgcggc accgccgtca accaggacgg cgccagcaac 116400

ggcctgaccg ccccgaacgg cccttcccag cagcgcgtca tccgccaggc cctcgccaac 116460

gcgggcctga ccgccgccga ggtggacgcg gtcgaggcac acggcacggg cacccggctc 116520

ggcgacccga tcgaggcgca ggcgctgctc gcgacctacg gtcaggaccg ccccgagggc 116580

agccccctgt ggctgggctc catcaagtcc aacttcggtc acacgcaggc cgccgccggt 116640

gtcgccggga tcatcaagat ggtccaggcg atgcaccacg gggtgctgcc gaagaccctg 116700

cacgtcgacg cgccgtcccc gcacgtggac tggtcggcgg gcgcggtctc gctcctcacc 116760

gagcagatgg cctggcccga aaccggccgc ccgcgccgcg cgggtgtgtc gtcgttcggc 116820

atgagcggta cgaacgccca cgccatcatc gaactcgccc cggacgccgc caccccgagt 116880

gccgcccggc cggagccggc cccggccgcc ctcccgtgga acctctcggc ccgcaccccg 116940

gacgccctgc gcgcccaggg cgagcggctg ctgtcccacc tggagaccca ctgtgagacc 117000

cacccggaga cggtgctcgc cgacatcggc cactcgctga cgaccggccg tgccctcttc 117060

gagcaccgcg cgacggtggt ggcgggcgac cgcgacggct tccgcgccgg actggccgca 117120

ctcgccgaag gccggacggc ggcgggcctg atccagggct cgtcctcgac cggcggtcgc 117180

acggcgttcc tgttcacggg gcaggggagc cagcggctgg ggatggggcg cgagctgtac 117240

gaggcgtatc ccgttttcgc gcgggctctg gacgaggtgt gtgcccgtct ggaactgcct 117300

ctgcctctga aggatgtgct gttcggtact gacacgggtc tgctgaacga gaccgcgtac 117360

acccagccgg cgctgttcgc cgtcgaggtg gcgctgttcc ggctggtgga gagctggggc 117420

ctgaagccgg acttcctggc gggtcattcg attggtgaga tcgctgctgc gcatgtggcg 117480

ggggtgctct cgctggagga tgcctgtgct ctggtgtcgg ctcgcgggcg gttgatgggt 117540

gcgctgcctg gtggtggcgt gatgatcgcg gtgcaggcgt cggagggcga ggtcctgccg 117600

ctgctgaccg accgggtgag tatcgccgcg atcaacggtc cgcagtcggt cgtgatcgcg 117660

ggtgacgagg ccgacgcggt cgcgatcgtg gagtccttct cggaccgcaa gtccaagcgg 117720

ctcacggtga gccacgcgtt ccactcgccg cacatggacg gcatgttgga cgacttccgg 117780

gccgtggcgg aaggcctgtc ctacggggcc ccgcgcatcc cggtcgtttc gaacctcacc 117840

ggggccctgg tctcggatga gatgggttcg gcggacttct gggtccggca cgtccgtgag 117900

gccgttcgct tcctggatgg catccgcgcc ctggaggccg cgggcgtcac gacatacatc 117960

gagctgggcc ccgacggcat cctgtcggcg atggcccagg agtgcatcac cggcgagggt 118020

gcggccttcg cgccgtcct gcgggcggga cgcgacgagg ccgagacggt gctctccgcg 118080

ctcgcggcgg ctcacgtccg cggcgttccc gtcgactggc aggccttcta cgccccggcc 118140

ggagcacagc gcgtgccct gccgacgtac gccttccagc gctccgtcta ctggctggac 118200

gcgggccggg cacagggtga catcgcctcc gctggactcg gcgcgacgga ccatccgctg 118260

ctcagcgccg cggtcgaact gcccgactcg gacggtttcc tcttcaccgg ccgcctgtcg 118320

ctggccaccc acccgtggct cgccgaccac gcggtcctgg gctccgtact ccttccgggt 118380

acggctttcg tcgaactcgc gctgcgggcc ggtgaccagg tcggctgcga cctgatcgac 118440

gaactcactc tcgaagcacc gctggtgctg cccccgcacg gaggcgtcca gctgcggctc 118500

gccgtcgcgg ccgccgacgc gacgggtcgg cgcaccctgg cgttccactc ccggagcgag 118560

gacgcggacg ccgggacgcc gtggacccgt cacgcctccg gtgtactcgc ggtcggggcc 118620

gagcggactc cgcagagcct caccgagtgg ccgccgaccg gggccgaatc cgtaccggtg 118680

gacgggctgt acgagggcct ggccgaatcc ggcttcggat acggtccggt cttccagggc 118740

ctgcgtgccg cctggcggcg cgacggcgag tactacgccg aggtcgccct gcccgagggc 118800

acggaggacg aggccggacg cttcggcctc cacccggccc tgctcgacgc ggcgctgcac 118860

gcgctgggtc tgggcagcac ggacaccgaa ggcggcgaag gacggctgcc gttctcctgg 118920

tccggtgtgc acctgcacgc cgtcggtgcc tccgcgctgc gcgtacgtct caccacgtcc 118980

cgaagcggtg aggtggcgct gaccatcgcc gacgcggccg gagagccggt cgcgaccgtg 119040

gccggcctcg cgctgcgggc cgtgagccgc gagcagctga gcacggcacg ggacctcacg 119100

cgtgacgcgc tgttccgggt ggactggact gcgttgcctg cgggcggtgc cgtggggtcg 119160

ctggacgact ggatgttgtt gggtgcgggt tcgcaggtgt atgcggatct ggcggggctg 119220

ggtgtggctg ttgcggaggg tggtgggatt ccggcggcgt tggtggtgcc ggtttcggag 119280

cctgatgcgg agtctgctgc gggtggtgtg gcgggtacgg tgcacgcggc tgttgagcgt 119340

gcgctgtctc tggtgcagga gtggttgtcg gacgagcggt tcgcggatgc gcgtctggtg 119400

ttcctgacgc ggggtgcggt ggctgcgcgg gccggggaca cggttccggg gctggtgcag 119460

gccgctgtgt ggggtctggt gcgctcggcg cagtcggaga atccgggtcg tttcgctctg 119520

atcgatgtcg acggcgacgg cgacggtgac ggtgaagtgg acggggacgt gctgtcggcc 119580

gcgctcgcca ccggtgagcc tgagctggcg gtccgtgaag gggctttgct cgtgccgcgc 119640

cttgcccgcg ccgctgtcgt tgagggtgcc ggtcgtgaac tggatgtcga cggcaccgtg 119700

ttggtcacgg gtgcgagcgg caccctgggt ggcttgttcg cccgtcatct ggtggttgag 119760

cgtggtgtgc ggcggctgct gttggtcagt cgtcgtggcg aggctgcgga aggtgctgct 119820

gaactgggcg ccgaactcac ggagctgggt gctgatgtgc ggtgggcggc gtgtgatgtg 119880

gccgaccgcg atgcgcttga ggctgtcctg gccgggattc ctgctgagta tccgttgtcg 119940

ggtgtggtgc atacggctgg tgtgctggac gacggtgtgg tgtcgtccct gaccccggag 120000

cgcctctcgg cggtgctgcg tccgaaggtg gatgcggcat ggaatctgca tgagctgacc 120060

cgcggtttgg atctgtcgct gttcgtgttg ttctcttcgg ctgccggagt gttcggcggt 120120

gcgggtcagg cgaactatgc ggcggcgaat gtgttcctgg acgctctggc ccagcaccgc 120180

agggcccagg gcctggccgc gacctccctt gcctggggtc tgtgggccgg tgtgggcggc 120240

atgggcggtg agctgacgga atccgaccgc gagcgcatca accgcggcgg catcaccgct 120300

cttgagcccg agaccggtct cgccctcttc gacgcggcac agcgcaccac cgacgcactg 120360

ctcgtccccc tcccgctcga cctggccgcc ctgcgcgtcc aggccggcag cggaatgctt 120420

ccggacctgc tgcgcggcct ggtccgcgta ccggtgcgcc gggcggcggg gcagggaagc 120480

gcggccgggg gcgggtcggt actccgtacc cgactggctg cgatgcccgc cgatgagcgg 120540

gacgcggccc tgctggacct ggtccgggcc gaggtggcgg ccgtactcgg ccacgcgtcg 120600

accgacgagg taccggccga ccgggcgttc aaggagctcg gcttcgactc gctgacctcg 120660

gtcgagctgc gcaaccgcct cggcgccacc acgggtgaac ggctctccgc caccctcgtc 120720

ttcgactacc cgaccccgca cgcgctcgcc gagttcctgc gcaccgaggt gctgggcctg 120780

gacgagccga cggatacggc cacgaccgcc cccacgcacc tcgggacatc gctcgacgac 120840

gacccgatcg cgatcgtcgg catgagctgc cggtaccccg gcggggtcga gaccccgag 120900

gacctctggc gcctggtggt gggtggcggc gacgccatct cggagttccc gcagggacgc 120960

ggctgggacc ttgagtcgct ctacgacccg gacccggacg gcaagggcac cagctacacc 121020

cggtcgggtg gcttcctgca cgacgcgggc cggttcgacc cggcgttctt cgggatctcg 121080

ccgcgcgagg ccgtggcgat ggacccgcag cagcggctgc tcctcgaaac ctcgtgggag 121140

gcgttcgagc gggccgggat cgacccggcc tcgatgcgcg gcagccggac cggtgtcttc 121200

gcgggcatca tgtaccacga ctacgcgacc cggatcacct ccgttccgga cggggtcgag 121260

ggctacctcg gcaccggaaa ctccggcagc atcgcctccg gccgcgtctc gtacgccttc 121320

ggcctggagg gcccggcggt caccgtcgac acggcctgct cgtcctcgct cgtcgccctg 121380

cactgggcga tccaggcgct gcgcaacggc gagtgcacga tggcgctggc cggcggtgtc 121440

accgtcatgt cgacgccggg caccttcacc gagttcagcc gccagcgcgg cctggccgcc 121500

gacggccgca tcaagtcctt cgcggccgcg gccgacggca ccagctgggc cgaaggcgcg 121560

ggcatgctgc tcgtagagcg gctgtcggag gcgcgggcca agggccaccc ggtcctggcg 121620

atcgtgcggg gctcggcgat caaccaggac ggtgcgagca acggcctgac cgctccgaac 121680

ggtccctcgc agcagcgggt gatccgccag gccctcgcgg gggcccggct gaccagtgac 121740

cagatcgacg tggtggaggc gcacggcacg ggcaccaccc tcggcgaccc gatcgaggcg 121800

caggcgctcc tggccacgta cggccgcgag cgcgaggcgg accagccgct gtggctgggc 121860

tcgatcaagt ccaacatggg tcacacgcag gcggccgccg gtgtcgcggg catcatcaag 121920

atgatcatgg ccatccggca cggtgtgctg ccgaagaccc tgcacgtcga cgagccgact 121980

ccgcatgtgg actgggaggc cggtgcggtc tcgctcctca ccgagtccgt cccgtggccg 122040

gagacgggcc gtccgcgccg cgccggtgtg tcgtcgttcg gtatcagcgg caccaacgcg 122100

cacacgatca tcgagcaggc gccggaggag ttcgtcccgg tccgtgtgac cgagtcgcag 122160

acgccgggcg cgggttcgcg agtgctgccg ttcgtgttgt ccgcgaagtc ggcgggggcg 122220

ttgcgtggtc aggcggtgcg tctgaaggcg catgtggagg cttcgccgga ggtgtctgga 122280

gccggggccg ttgatgtggc gtattcgctg gcgacgcggc gtgcggtctt cgaccaccgt 122340

gcggtggtgg tggccggtga ccgcgaggag ttgctgcgtt ctctggctgc tgtggagtcg 122400

gagggcgcgg cggctggtgt gacccgtggg gccgtgggtg gcggaaagct tgccttcctg 122460

ttcacgggcc aggggagcca gcggctcggg atgggccgtg agctgtacga gacgtatccc 122520

gtcttcgcgc gggctctgga cgcggcgtgt gctcgtcttg aactgccgct gaaggatgcg 122580

ctgttcggca ccgatgcggg tctgctgggc gagacggcgt acacccagcc ggctctcttc 122640

gcggtcgagg tggcgttgtt ccgactgctg gagagctggg gtgtgaggcc ggacttcctg 122700

gcgggtcatt cgatcggtga gatcgcggcc gcccatgtgg ccggggtgct ctccctcgat 122760

gacgcctgcg cactggtcga ggcgcgtggt cgtctgatgc aggcgctgcc gaccggtggc 122820

gtgatgatcg ccgtccaggc gtctgaggct gaagtcctgc cgctgctgac cgaccgcgtg 122880

agtatcgccg cgatcaacgg tccgcagtcg gtcgtgatcg cgggtgacga ggccgacgcg 122940

gtggcgatcg tggagtcctt ctcgggccgc aagtccaagc ggctcacggt cagtcacgcg 123000

ttccactcgc cgcacatgga cggcatgctg gctggcttcc gcaaggtggc ggagagcctg 123060

tcgtacgagg ctccgcgcat cccggtcgtc tcgaacctca ccggggccct ggtcaccgac 123120

gagatgggtt cggccgactt ctgggtgcgg cacgtccgcg aggccgtccg cttcctggac 123180

ggtatccgca ccctggaagc cgcaggcgtc gcgacgtacg tcgaactcgg ccccgatggc 123240

gtcctgtcgg cgatggccca ggactgcgtc accggcgagg gtgcggcctt cgcgcccgcc 123300

ctccgcaagg gccgccccga gaccgagacg atcaccacgg ccctcgccct tgcccacgcc 123360

cacggcacgt ccgtcgactg ggagacgtac ttcgccggga ccggcgccca gggcgtcgag 123420

ctgccgacct acgccttcca gcgtgactgg tactggctga actcggccgt ggtgcaggcc 123480

ggtccgggcg acgcgagcgg attcgggctc ggcgcgaccg atcacccgct gctcgacgcg 123540

accatcgaac tgcccgactc ggacggcttc ctgttcacca gcaggctgtc cctcgacacg 123600

cagccgtggc tcgcggacca cgccgtcctg gggtcggtcc tcctcccggg cacggccttc 123660

gtggaaatcg ccgtacgggc aggtgaccag gtcggttgcg acgtactgga agagctgacg 123720

ctggaggcac cgctggtggt gcccgagcgg ggcggtgtgc agctgcggct caccgtcgcc 123780

gccgccgacg agtcgggacg gcgaggtctg tcgctgtact cccgcgacga ggacgctccc 123840

gccgacgagc cgtggacgcg ccacgccagc ggcgtgctcg ccaccggcgc ggcggccccc 123900

gacttcgacc tcgccgcctg gcccccggcc ggagccgaac cggtcgacat cgacggcctg 123960

tacgagggcc tggccgcggc cgggttcgac tacggtccgg ccttccaggg cctgcgcacg 124020

gcatggctgc acggcgacgc ggtgtacgcc gaggtgagcc tggacgagga gtccgcggaa 124080

tcggcggaat ggttcgggct gcacccggcc ctcctggacg cgacgctgca cgcggcgggt 124140

ctcggcggtc tcgtggagag caccggccag ggacggcttc cgttcgcctg gagcaatgtg 124200

tccctgcacg cggccggcgc gtccgcggta cgggtccggc tggccccggc cggccgtgac 124260

gcggtgtctc tgcagctcgc cgacgcggcg ggcgcaccgg tcgcctcggt cgaatcgctg 124320

gtgctgcggg cggtctcgcc cgaccagatc ggcgcggcgc gcggcggccg tcacgagtcg 124380

ctcttcgaga tcgactgggc cgccctcccg ctcgccccgg tgtccgctgc cgaacagcgc 124440

ccctgggcgc tgctggcgga cgacgggtcc ggccacgcgg gactcgaagc cgtgggtgtc 124500

cgtcacgagg cccacaccgg actcgcggcg ctcgccgaca ccggacgggc gatccccgag 124560

gtcgtgtgcg tcccgctcgc tgcggcgaac tcccaggacc tggcgggtgc gggtgcggtg 124620

cacgcggctg tggagcgtgc gctgggtctg gtgcaggagt ggttgtcgga cgagcggttc 124680

gcggatgcgc gtctggtgtt cctgacgcgc ggtgcggtgt ccgcggtgcc gggcgaggac 124740

gtgaccgatc tggtccacgc tccggtgtgg ggtctggtgc gttccgcgca gtccgagaac 124800

ccgggccgct tcgtcctggc cgacaccgac ggcaccgacg cctcctaccg tgccctgacg 124860

gccgcgctcg cctcgggcga gccggagttc acggtgcggg gcggcgcggt acgggtgccc 124920

aggctgacgc gctccactgc tgtcgctgtg gaggctgtgc ccgaactcgg ttcggacggc 124980

acggtgttgg tgacgggtgc gagtggcacg ttgggtggtt tgttcgcccg ccatttggtg 125040

gttgagcgtg gtgtgcggcg cctgctgttg gtgagtcgtc gtggtggggc tgcggagggt 125100

gctgctgaac tgggcgccga actcacggag ctgggtgctg atgtgcggtg ggcggcgtgt 125160

gatgtggccg accgtgatgc gcttgagtcc gtcctggccg ggattcctgc tgagtatccg 125220

ttgtcgggtg tggtgcatac ggctggtgtg ctggacgacg gtgtggtgtc gtccctgacc 125280

ccggagcgcc tctcggcggt gctgcgtccg aaggtggatg cggcatggaa cctgcacgag 125340

ctgacccgcg gtttggatct gtcgttcttc ctgttgttct cgtcggctgc cggtgtgttc 125400

ggtggtgccg gtcaggcgaa ctatgcggcg gcgaatgtgt tcctggacgc tctggcccag 125460

caccgcaggg cccagggcct ggccgcgacc tcccttgcgt ggggtctgtg ggctgagccg 125520

gggggcatgg cgggcgcgct ggacgctgat gatgtgtcgc gtctgggccg tggcggtgtc 125580

agcgggctct ccgcgcagga gggtgtggcg ttgttcgacg cggcgtccgc ctccgaacag 125640

gccctgttcg ttcccgtgaa gctggacctg gccgccctgc gcgcccaggc gggtagcggc 125700

atgcttccgc cgctgctcag cggtctcgtc cgtaccccca cccgccgcgc cgcgggcacc 125760

ggcggcaccg gagacaccgg cacggacggt gggaccgcgc tgcgggagcg cctggccggg 125820
ctcgcaccgg ccgcgcggga cgaagcgctg ctggagctcg tctgcacgta cgtcgcggcg 125880
gtgctcggct tcgccgggcc cgaggcggtc gatccggcgc ggtcgttcag cgaggtcggc 125940
ttcgactcgc tgaccgccgt cgagctgcgc aacaggctcg gcgccgcgac cggcgtacgc 126000
ctccccgcca ccctcgtctt cgactacccg acaccggacg cgctggtgga gtacctgcgc 126060
gacgaactct ggcaggacgg cgccgcggcg gtaccccccgc tgctcgccga actcgaccgg 126120
ctggagaaga cgctcgtggc gtccgtgccc gacgacgacg gccgcacccg catcaccgag 126180
cggctgcagg ccctgctggc cgcctggagc gaggccggcg aatcaacgga caccgccgac 126240
gccgatgtgg ccgaggcgct tgagaccgcg accgacgatg acctcttcga cttcatcggc 126300
aaggagttcg ggatctcgtg atgcgaaggc ccggctccgc cctttccgac ggctctgtct 126360
ttctggcttc tgtacgaggg atgcacgcat gaatgaggaa aaactccggt acttcctgaa 126420
gcgggtgacg gccgatctcc acgagacgcg ccggcgtctt caggaggtcg agtcggagga 126480
gcaggagccg atcgcgatcg tcgggatgag ctgccgctac ccgggagacg tcgagtcgcc 126540
cgaggacctg tggcggctgg tgtccgagga gaccgacgcc atctcccctt tccccaccga 126600
ccggggctgg gacatggggc ggctcttcga cgcggacccc gacgggcggg gcacgagcta 126660
tgtgcaggaa ggcggcttcc tgcactccgc caaccggttc gacccggcgt tcttcgggat 126720
ctcgccgcgc gaggccgtgg cgatggaccc gcagcagcgg ctgctcctcg aaacctcgtg 126780
ggaggcgttc gagcgggccg ggatcgaccc gacctcgctg cgcggcagcc ggaccggcgt 126840
cttcgcgggc gtcatgtacc acgactacgc ctcgcggctg cgtgccgtcc cggaggaggt 126900
cgagggttac ctcggcaccg gcggctccag cagcatcgcc tccggccggg tctcgtacac 126960
cttcggcctg gagggcccgg cgctcaccgt cgacacggcc tgctcgtcct ccctcgtcac 127020
gctgcacctg gccatgcagg cgctccgcaa gggcgagtgc tcgctcgccc tcgcgggcgg 127080
tgtcaccgtg atggcgacac cgggcacctt cacggagttc agccgccagc gcggtctgtc 127140
cttcgacggc cgctgcaagt ccttcgcgga ctccgcggac ggcaccggct gggccgaggg 127200
cgcgggcatg ctcctcgtgg agcggctctc ggacgcccgt aagaacggcc atacggtact 127260
cgccgtggtc cggggctcgg ccgtcaacca ggacggtgcc agcaacggcc tgaccgcccc 127320
gaacggcccc tcccagcagc gggtcatccg gcaggccctg gccgacgccc gcctcacggc 127380
ggccgacgtc gacgtcgtgg aggcacacgg caccggcacc accctcggtg acccgatcga 127440
ggcgcaggcc ctgctcgcca cgtacggccg ggaacacacc gaggacagcc cgctgtggct 127500
cggctcggtc aagtcgaacc tcggtcacac ccaggcggcc gcgggcgtcg ccggcatcat 127560
caagatggtc atggcgatcc gccacggccg gatccccaag acgctgcatg tcgacgagcc 127620

gtcgaccaac gtcgactggt cggcgggcgc cgtctcgctg ctgcgggagt ccgtggagtg 127680

gccggagacc ggccgcccgc gccgcgcggc gatctcttcc ttcggcatca gcggcactaa 127740

tgcgcacacg atcatcgagc aggctccgct gccggaggcc gagaccgaaa ccgagccgac 127800

cggcgacgag acggacggct ctgagagcac ggcgggggca gaggggacag aggggacaga 127860

gggcgccggg gtgcggcccg tgtccgtgcc tcccgtcctt ccgtggcccg tctcggcccg 127920

tacggaggag gccctgcacg cccaggcgga acgcctgctg gcccacgtgc ggaccaaccc 127980

ggaccaggcc ccggtgggcg tcgctctctc cctggccaca gggcgcgccg cgctggaaca 128040

ccgcgccgtt gtcgtcgcca ccgaccggga aaccgccctc gccgacctcg ccgcactggc 128100

gtccggcgag acctcggcgc gcgtcgtgct cggcgagccg ggagcgcggg gcaagaccgc 128160

gttcctgttc acggggcagg ggagtcagcg gctggggatg gggcgcgagc tgtacgagga 128220

gtatcccgtc ttcgcggatg cgctggacgc ggtgtgtgcc cgtcttgaac tgcctctgaa 128280

ggatgtgttg ttcggggcgg atgcgcgtct gctggacgag accgcttata cgcaaccggc 128340

gctcttcgcc gttgaggtgg cgttgttccg gttggtggag agctggggtc tgaagcccga 128400

cttcctggcc gggcattcga tcggcgagat cgccgccgcg cacgtcgcgg gggtgttctc 128460

gctggaggat gcttgcgcgc tggtgtcggc tcgtggccgg ttgatgggtg ccctgcctgc 128520

gggtggcgtg atgatcgcgg tgcaggcgtc ggaggacgag gttctgccgc tgctgacggc 128580

ccgggtgagc attgccgcga tcaatggtcc gcagtcggtg gtgatcgcgg gtgacgaggc 128640

cgacgcggtc gcgatcgtgg agtccttcac ggggcgtaag tcgaagcggc ttacggtcag 128700

tcacgcgttc cattcgccgc acatggacgg gatgttggaa gacttccggg tcgtggcgga 128760

ggggctgtcg tacgaggctc cgcgcatccc cgtcgtttcg aacctcaccg gggccctggt 128820

ctcggatgag atgggttcgg cggacttctg ggtccggcac gtccgtgagg ccgttcgctt 128880

cctggatggc atccgggccc tggaggccgc gggcgtcacg acgtacgtcg aactcggccc 128940

cgacggtgtc ctgtcggcga tggcccaggc atgcgtgacc ggcgagaact ccgtcttcgt 129000

gccggtcctg cgctcgggtc gctccgaggc ggagagcgtc accacggccc ttgcccaggc 129060

gcatgtccgc gggatcgccg tggactggca ggcctacttc gccggtaccg gtgccgagcg 129120

cgtcgacctg cccacctacg ccttccagcg cgaccactac tggctcgacg ccggaacgct 129180

cggcggagac gtgaccacgg cgggccttcg atccgccgat caccctctgc tcggcgcctc 129240

tgtggctctg gcggatgcgg agggccttct cctcaccggc cggctctcgc tcgacaccca 129300

cccgtggctc gccgaccacg ctgtggcggg gacggtcctg ctgcccggta cggcgttcgt 129360

cgaactcgcg ctgcgggccg gtgaccaggt cggctgcgac ctgatcgacg aactcaccct 129420

cgcggcgccg ctggtgctgc ccgagcaggg tggagtcgaa ctccagatca ccgtcgcggc 129480

ccccgacgaa tcgggccgcc ggtccgtcgc cttccactcg cgcgccgaca gcgccgcgga 129540

cgacgaggcg tgggtccggc acgcgaccgc agtactggcc gagggcgcgg acaccccggt 129600

gttcgacttc ggcgtctggc cgccgaccgg ggctgaatcc gtaccggtgg acgggctcta 129660

cgaggggctc gcgcactccg gattcggcta cggtcccgtg ttccaggggc tgcgtgccgc 129720

ctggcgccag ggcgaggacg tgttcgccga agtgagcctc ggggacgggg tcgagcccgg 129780

agcagcgcac ttcaccgtgc acccggccct gctcgactcc gccctgcacg ccatcaacct 129840

cggcaccctc gtcgaggaca ccggccaggg gcgactgccg ttcgcatgga gcggggtcgc 129900

ggttcacgcc gtgggggcgg acaccctgcg cgtacggctc tcccgggccg gtcaggacgc 129960

ggtggccctg gagatcgcgg acgcggacgg cgcgcccgtc gcttccgtac gcagcctggc 130020

cctgcgcgcc ttctcacccg accagctgac cgggccggac ggcgccggtc acggcgacgc 130080

gctgttccgg gtggactggg cggcgttgcc tgcgggcggt gcggtcgggt cgctggacga 130140

ctggatgttg ttgggtgctg gttcgcaggt gtatgcggat ctggcggggt tgggtgtggc 130200

tgttgcggag ggtggtggga ttccggcggc gttggtggtg ccggtttcgg agcctgatgc 130260

ggagtctgct gcgggtggtg tggcgggtgc ggtgcatgcg gctgttgagc gtgcgctggg 130320

tctggtgcag gagtggttgt cggatgagcg gttcgcggat gcgcgtctgg tgttcttgac 130380

gcggggtgcg gcggctgcgc gggccgggga cacggttccc gggctggtgc aggcggccgt 130440

gcggggtctg gtgcgctcgg cgcagtcgga gaacccgggc cgtttcgctc tgatcgatgt 130500

cgacggcgat ggtgaagtgg atgcggaggt gctgtcggcc gcgcttgcta cgggtgagcc 130560

cgagctggca gtccgtgaag cggctttgct cgtgccgcgc cttgcccgtg ccgctgtcgc 130620

ggtggagcct gcgcccgaac tcggttcgga tggcacggtg ttggtgacgg gtgcgagtgg 130680

cacgttgggt ggtttgttcg cccggcattt ggtggttgag cgtggtgtgc ggcggctgct 130740

gttggtcagt cgtcgtggtg aggctgcgga aggtgctgct gaactgggcg ccgaactgac 130800

tgggttgggt gctgatgtgc ggtgggcggc gtgtgatgtg gccgaccgtg aggcgcttga 130860

gtcggtcctg gccgggattc ctgccgagta tccgttgtcg ggtgtggtgc ataccgctgg 130920

tgtgctcgat gacggtgtgg tgtcgtcgct gactgccgag cgtgtgtcgg cggtactgcg 130980

tccgaaggtg gacgcggcgt ggaacctgca cgagctgacc cgtggcctgg atctctcgct 131040

cttcgtgttg ttctcgtcgg ctgccggtgt gttcggtggt gccggtcagg cgaactatgc 131100

ggcggcgaat gtgtttctgg acgctctggc ccagcaccgc agggcccagg gtctggccgc 131160

gacctctctt gcgtggggtc tgtgggatga gccgggggcc atggcgggcg cgctggacgc 131220

tgatgatgtg tcgcgtctgg gccgtggtgg tgtcagcgga ctctccgcgg gggagggtgt 131280

ggcgttgttc gacgctgcgt ccgcgtccga acaggccttg ttcgttccgg tgaagctgga 131340

cctggccgcc ctgcgtgccc aggcgggcag tgggatgttg ccgccgctgc tcagcggtct 131400

tgtccgtacc cccacccgcc gcgccgcccg gggcggttcg gccgcggggg gaacgttcgc 131460

ccggaagctg gccggcctcg cggtggacca gcggtccgca gccgtgatgg agctcgtgcg 131520

tgctcaggtc gcagccgtgc tcggccttgc cgggcccgaa gcggtagacc cggcacggtc 131580

gttcagcgag gtcggcttcg actcgctgac cgccgtcgag ctgcgcaaca ggctcggcgc 131640

cgcgaccggt gtacgcctcc ccgccaccct cgtcttcgac tacccgacct ccctcgccct 131700

cgccgacttc ctgggtggcg aactgctcgg cggtcaggaa gcggcagcag ccccgacggc 131760

cttcacggcc cgggacgacg agccgatcgc gatcgtggcg atgtcttgcc gtttccccgg 131820

cggcgtgcgg tcgcccgagg atctgtgggg gctggtcctg gacggccggg atgccatctc 131880

ggacatgccg gacgaccgcg gctgggacgt cgagggactc ttcgaccccg accccgaccg 131940

cccgggcacc agctacagca gggcgggcgg gttcctgcac gacgcccacc acttcgaccc 132000

gacgttcttc gggatctcgc cgcgcgaggc cctcgccacc gacccccagc agcggctgct 132060

cctcgaaacc tcgtgggagg cgttcgagcg ggccgggatc gatccggcca ccgtacgcgg 132120

cagccggacc ggcgtcttcg cgggcgtcat gtacaacgac tacggcaccc tcctgcaccg 132180

cgccccggag ggcctcgaag gctatatggg cacctccagc tcgggcagcg tcgcctcggg 132240

ccgggtctcg tacaccttcg gtctggaggg cccggcggtc accgtcgaca cggcctgctc 132300

gtcctcgctc gtcaccctgc acctcgccgt gcaggccctg cgcaacggcg agtgcgacct 132360

cgcgctggcc ggcggtgtca cggtgatggc cacgcccggt acgttcgtcg cgttcagccg 132420

tcagcgcggc ctcgcgagtg acggccgctg caagccgttc gccgcggccg ccgacggtac 132480

ggcgtggggc gagggcgtcg gcatgctgct cgtcgagcgc ctgtcggacg ctcgggccaa 132540

gggccacccg gtgctcgcgg tggtccgtgg ctcggcgatc aaccaggacg gtgccagcaa 132600

tggcctgacg gctccgaacg gtccctcgca gcagcgggtg atccgccagg cgctggccag 132660

tgccggtctg tcggcggcgg atgtggacgt agtggaggcg cacggcaccg gcaccaccct 132720

gggcgacccg atcgaggcgc aggcactcct cgccacctac ggtcaggagc acacggacga 132780

cagcccgctg tggctggggt ccatcaagtc caacttcggt cacacgcagg ccgctgccgg 132840

tgtcgcgggc atcatcaaga tggtgcaggc gatgcaccac ggggtcgtcc ccaagacgct 132900

gcacgtggac gagccgtccc cgcacgtgga ctggtcggcg ggcgcggtct cgctcctcac 132960

cgagcagatg gcctggcccg aaaccggccg tccccgccgc gcggcgattt cttccttcgg 133020

tatcagcggt accaacgcgc acacgatcat cgagcaggcg ccggaggagt tcgctccggt 133080

ccgtccggtc cgtgtgatcg agccggaggc ggtgggtgcg ggttcgcggg tgctgccgtt 133140

cgtgttgtcc gcgaagtcgg cgggggcgtt gcgtggtcag gcggtgcgtc tgaaggcgca 133200

tgtggaggct tcgccggagg tgtcgggggc cggggctgct gatgtggcgt attcgctggc 133260

gacgcggcgt gcggtcttcg accaccgtgc ggtggtggtg gccggtgacc gtgaggagct 133320

gttgcgtgct ctggctgctg tggagtcgga gggcacggcg gctggtgtga cccgtgggac 133380
ggcgggtggc ggaaagcttg ccttcctgtt cacgggccag gggagccagc ggctggggat 133440
ggggcgtgag ctgtacgaga cctatcccgt cttcgcgcgg gctctggacg cggcgtgtgc 133500
tggtctcgaa ctgccgctga aggatgcgct gttcggcgcc gatgcgggtc tgctggacga 133560
gacggcgtac acccagcccg ctctctttcgc ggtcgaggtg gcgttgttcc gactgctgga 133620
gagctggggt gtgaggccgg acttcctggc cgggcactcg atcggtgaga tcgcggccgc 133680
gcatgtggcc ggggtgctgt ccctggacga cgcctgtgcg ctggtcgcgg cccgcggccg 133740
gctcatgcag gcgctgccca ccggcggtgt gatgatcgcc gtccaggcgt cggaggacga 133800
ggtcctgccg ctgctgaccg accgggtgag catcgccgcg atcaacggtc cgcagtcggt 133860
cgtgatcgcg ggcgacgagg ccgacgcggt ggcgatcgtg gagtccttct cgggccgcaa 133920
gtccaagcgg ctcacggtca gtcatgcgtt ccactcgccg cacatggacg gcatgctggc 133980
tggcttccgc aaggtggcgg agagcctgtc gtacgaggct ccgcgcatcc cggtcgtctc 134040
gaacctcacc ggggccctgg tcaccgacga gatgggttcg gccgacttct gggtccggca 134100
cgttcgcgag gcggtccgtt tcctggacgg tatccgggcc ctggaggccg cgggcgtgac 134160
ggcgtacgtc gaactcggtc ccgacggtgt tctgtcggcg ttggcccagg agtgcgtcac 134220
cggcgagggt gcggccttcg cgcccgccct ccgcaagggc cgccccgagg ccgagacgat 134280
cacaacggcc ctcgcccttg cccacaacca cggcacgtcc gtcgactggg agacgtactt 134340
ctccgggacc ggcgcccagc gcgtcgacct gcccacctac gccttccagc gcgagcgcta 134400
ctggatcgac gtgcccgtcc actccgtcgg cgacgtggcc tccgccggac tcggtgcggc 134460
ggagcacccg ctgctgggcg cggccgtcga actgcccgac tccgacgggc tgctgctcac 134520
cggtcggctg tcgctcctgt cgcaccctg gctggccgat cacgccgtcg cgggcaccgt 134580
tctgctcccc gggaccgcct tcgtggagct ggcgctccac gccgggcagc gggtgggcag 134640
tggcctgctc gaagagctga ccctggaggc gccgctggtg cttcccgagc gcggggcgct 134700
ccagctgcgg gtgtccgtgg ccgcgcccga cgaggcgggg cgtcgtgcgc tgcacgtgca 134760
ctcgcgtccc gaggacctgg gcggcgagga ccgtacgggg cacgaggtgc cgtggacgcg 134820
gcacgccggc ggtgtgctcg ccgcgccgga ggcggccggt gccgcgccgg aggagtccgg 134880
cctggacgtc tggccgcccg cggacgccga accgctcgat gccggcgacc tgtacgaccg 134940
gttcgccgag ggcgggttcg cgtacggtcc tgtcttccgc aacctgcgcg ctgcctggcg 135000
gcgcggcgac gagctgttcg ccgaactgct cctgcccgag gggcagctcg cccaggccgg 135060
ccacttcggt gtgcacccgg cgctgctgga cgcgggtctg cacggcctcg cgctcggctc 135120
gttccatgac ggtgcggacg aggacgcccg gatccggctc ccgttctcct tcagcggtgt 135180

cgctctgcac tcggtcggcg cgggctcgtt gcgcgtacgg ctcgccccgg ccgggtccgg 135240

cgcggtgtcg ctcgcggcct tcgacgagca gggcgcaccg gtcgtgtcgg tggaatcact 135300

gctgctgcgg gcggtggatc cggcacggct gaaggccgcg gaacagccgg tgttccacga 135360

gtcgctcttc cggctggagt ggccggcgct ggccgcgggc ccgcgtacgg acaacgcccc 135420

cggggacggc ggccggtggg ccgtggtcgg ggccgactcg ctcggccttg aggccgggct 135480

gcgggcggac ggcgtcgccg tcgacgggta cgcggacctg tccgcgctcg ccggagtcgt 135540

ggccgcgggc aagccgcagc cggacacggt gctggtctcg tacgcctcct cgggtcccgg 135600

catcaggacg gcggacgccg ttcggcaggc ggctcacgac gcgctggagc tggtccaggg 135660

ctggctcgcc gaggagtcgc tcgccgggtc acgactggtc gtggtcaccc gcggcgcggt 135720

cgaggcgcgg cccggcgagg gcgtgcccga tctggcgcac gcggcggtgt ggggcctgct 135780

gcggtccgcg cagtccgaga accccgggcg gttcgtactg ctcgacctcg acgcggaaga 135840

cgcggaggtc ctggctccgc tgatggccgc cgctgtggcg agcggggaac cccagctcgc 135900

cgcccgcgag ggcgtcctgc atgccgcgag gctggcacgg gttcccgccg cccccaccgc 135960

ggtggcgggc acggagcgcg cgcccgccct cgaccccgac ggtacggtcc tcatcaccgg 136020

cggcaccgga tcgctcggca gcctgctggc ccgccacctg gtcgtggagc acggcgtacg 136080

gcacctgctg ctgaccagcc ggcgcggtgc cgccgccgag ggcgccccgg aactcgtcgc 136140

cgcactggcc gaactgggcg ccgaggcgac cgtcgccgcg tgtgacgccg ccgaccggga 136200

ggcgctggcc gcgctgctgg ccggcattcc ggccgcgcac cccctcacgg ccgtcgtcca 136260

cacggcgggc cgcgtcgacg acgggctcct ggcgtcgctc agcccggagc ggatcgacac 136320

ggtgctgcgt cccaaggccg acgcggcgct gcatctgcac gagctgaccc gcgggctgga 136380

cctcgccgcg ttcgtcctgt tctcctccgc ggccggaacc ctcggcaacc ccggccaggc 136440

caactacgcg gcggccaacg ccttcctgga cgccctggca cagcaccggc gcgcggcggg 136500

gctgcccgcg gtgtcgctgg cctgggggct gtgggagcag cgcagcgcga tgaccggagc 136560

gctgtcggac gcggacgtcc agcggatggc acgcgccgga ctcgcgcccc tctcctcggc 136620

ggagggcctg gccctcttcg acacggcgtg cgccctcgcg ccggtgggcg ccacggagac 136680

cgccaccggc gacggagcgt tcgtcgccat gcggctggac accgcgcccc tgcgggccca 136740

ggcggacgcc ggagcccttc cggcggtctt ccgcgggctg gtgcgcggag gtcctcgcag 136800

ggccgccgca catcaggccg ccgattcggc ggcatccact gccgcgcgaa agctcgcggg 136860

cctgtccggg ctgccgcagg acgagcagga gcgcgtgctg ctcgacctgg tgcgcgccca 136920

ggtggccgcc gtactcgcct atccgtcgcc ggacgcggtg ggggagtcgc aggagttcct 136980

ggagctgggt ctggactcgc tgaccgccgt cgagctgcgc aaccagctga acgcggcgac 137040

cggcctgcgg ctgcccgcca ccctgctctt cgaccacccc actcccgcgc tggtcgccga 137100

gcggctgcgc gccgaactcg ccggagcctc cggcccggcg gcggtccggg agggcgcggc 137160

ggacagcggc gcggagggct ccgcgggtgt cttcggggcc atgctccacg aggccggaac 137220

gcagggtgcg tccgggcagt tcatggagct gctcatgcag gcgtcgcggt tccggccgtc 137280

gttcgcctcg gcggccgagc tgcgcaaggc gccgagcctc gtgcggctct cccgcggtga 137340

cacccggccg ggactggtct gtttctcctc gatcctgtcg atctcgggcc cgcaccagta 137400

cgcgcgcttc gcctccgcgt tccggggccg ccgggacgtg cacgcgctcg gtgcccccgg 137460

cttcctgcgg ggcgagcagc tgccctcggc caccgacgcg gtgatcgagg cccaggcgga 137520

ggccgtgctc cggcacgcgg acggtgcgcc gttcgtcctc ctcggccact cctcgggcgg 137580

catgctcgcc cacgcggtgg ccgggaggct ggagagcgag ggggtcttcc cccaggcgct 137640

ggtgatgatc gacatctact cgcacgacga cgacgcgatc atcggcatcc agcccggcct 137700

ctccgagggg atggacgagc ggcaggacac ctacgtaccg gtcgacgaca accggctgct 137760

ggcgatgggc gcgtacttcc ggctgttcgg aggctggaag cccgaggtgg tgaagacgcc 137820

gaccctgctg gtccgggcgg gtgagcggtt cttcgactgg acccggtcca cggacggcga 137880

ctggcgttcg tactgggacc tggaccacac ggccctggac gtgccgggca accacttcac 137940

catgatggag gagcacgctc cgacgaccgc acaggccgtc gaggggtggc tggacacgac 138000

cggctgacac caccggctga cggcgccgga cagcgacatg gccgggcgtc aagcgtcaga 138060

cgtcaggcga cgcgcttctc acgctcgcgg gagcgcttct tcggcagccc caccgtcacg 138120

acctcgaagc tgtccttggt gaggtcgagg cggtggaaga ggttgtcggg cccggtcacg 138180

cacaccgtgc ccacgccgag cccccttgagg gactccacca cgcccggcca gtggacgggc 138240

cggtcgaagg tgtccagcat catcgtgcgc atcccggcgg cgtcccggac gaccccgccg 138300

tcctggtcgt tgaccacggg cagggtgggg tcggccagtt cgtacgcggc gaagacctct 138360

tcctccgcct tgcggcgcag cgccgagaag gccgccgcgt gcacgggcgg gcgcatcgag 138420

tacatggagt agccgccgac cgcgctgatg cccgccttca gcccgtccag ctccttctcc 138480

tgtacggaca ccatgtggaa agcggcgtcc agccgcccgg agatgtcgta ccaggcaccg 138540

cggtcgtcga agccggccag gatctcgtcc agccggtcct gcggggtgcg gacgaagcag 138600

tgcgtgacga cgtcctggta cgcgtcggcg aagtactcct cctcgcagcg ggccagctcc 138660

gcggtgagcc ggacgacgtc cgcgaagggc agcgacccga cgaaagcgga ggcggccttc 138720

tggccgaaac tcgggccggc gcagacggtg ggagagatgc cgagcgcgtc caccgcccgg 138780

tcggccatag ccatcgaatt caccaggaag gcgatctgcg aatagaccga gtagtcgtcc 138840

tcggaggtgc ggaaacggtc gaacaccgaa tatccgagcg cctcgtctgc ctccgcgagg 138900

cgccggcgcg cgtaagggtc gagcagcagg aactttccga cctccgcgaa ggacgagggg 138960

cccataccgg gaaagacgat cgccgtctcg gtcgagggag tctgctcgga gtcgaagccg 139020

gagttgaagc cggagtcgga gccggaacgg gagtcggaac gggaatcaga agtggtcatg 139080

atccgtgaat gcctttgctt ccggggacgg caccggcagg cacctgccgc cgtcacgaac 139140

gtaggaacgg ccccgcaccc ggccggacgc gaatgcgccg agccgggcac gaggccagga 139200

gggacgagag gggggagacg agagagggga gaccagacgg ggcagcgcgc gctcagtcct 139260

gcgcctcagt cctgcgccct gcggtggaac cccttgatgc cgatcagccc gaagaccacg 139320

atcgccccgc tcagggcgag cagatcgatc cacagcggaa tcgagccggg gccgcccggc 139380

ggcagcagca gggcgcggat cccctcgctg acgtaggtca gcgggttgat ggcgcacagc 139440

acctggaacc agcggatgtc cgccaggctg tgccagggga actgggtgca gccggtgaac 139500

atcagcgggg tcagcgtcac ggcgaagatg acgctgatgt gccgcggcgg ggccagcgtg 139560

ccgatggtca gacccaccgt gctgcccgcc agcgcgcccg tcagcagcac gcccagcgtg 139620

ggcaggaagc tgtccatcgg ccaggacacg tcgtcgagga tcaggaagcc gacggggatc 139680

atcaccagtg aggcgatgat gccgcgcagc gccccgaaga ccagcttctc gacggccacc 139740

aggctggtgg ggatgggcgc gaggagccgg tcctcgatct ccttggtcca ggagaagtcg 139800

atgaccaggg gcagcgcggt gttctgcagg ctgaccagga agctgttgag cgcgaccacg 139860

cccgggagca ggatctgctg gaacccgccg ccggtgtaac cgagttcgcc gaggaccttg 139920

ccgaagacga acaggatgaa gaacggttcc acgagcacct gggcgaggaa cgggcccagt 139980

tcgcggccgg tgacgaagat gtcccgccac aggatgaaga agaacgtgcg ggtcgcggtg 140040

cgcacgtcgg tgcgcgcggg ccgcagttcg gccgggaagt cggtgaccgg gtcgggtgcg 140100

gtcagggtgg ccgtcatcgc agctcccggc cggtgagctt gatgaagacg tcctccaggg 140160

tcgcggttcc gacgctcacg tccttgatgt cgtgactcgc ttccgtcagg gccgtgatgg 140220

cggtcggcag caccgcgccg gacggcgcgt cgctgtagag gcggagccgt accggcgcgg 140280

gcgcgccgcc ctgctccttg gcgtgttcct ggtgtgccag ctcgacccgc tcgaccgtct 140340

cgatccgctc cagcaggcgt acgacgctct cggcgtcgtg ccccgcgggc tggacggtga 140400

gggtgagggc ggtgctgctc aggctccggg tcagcgcctg cggggtgtcg agggccagca 140460

gtcggccgtg gtcgacgatg ccgacgcggt cgcagagctt ggcggcttcg tccatgtcgt 140520

gcgtggtcag cacggtggtc accccgcgct tgctcagctc ggccacgcgc tcgtggatga 140580

acagccgtgc ctgcggatcg agtccggtgg cgggctcgtc gaggaagagc acgtcggggc 140640

ggtgcatcag ggcccgggcg atcatcacgc gctgggcctg gccgccggag agttcgtcgc 140700

cgcgggcctt gccccggtcg gcgagaccca cccactccag gcactcgtcg gcgagccgtc 140760

cgcgttcgga gcggctcatg ccgtgatagc cggcgtggaa ggtcaggttc tgccggaggg 140820

tcagcgaccg gtcgaggttg ttgcgctgcg gtacgacggc gaaggcccgg cgcgcctggg 140880

cggggtgggc cacgacgtcg acgccctgga cgaacgctcg ccccgccgtg ggggccacgc 140940

gggtggtgag gatgccgatg gtcgtcgtct tgcccgcccc gttcgggccg aggaatccga 141000

agacctcgcc cctgcggacc gagaagctca ggtcgtccac cgctggtcgg tcgcggctcc 141060

ggtacttctt gactagtccg tcgaccacga cggcggaatc cacgggtcgt tcagagttca 141120

tttacgcctg cgaatcaagc gggacgcggc gacggcagtc cgggggattc gcacaggaat 141180

gtcgcgtgac cggccgcgcg tcgagcgccg actgaatagg gcataggagt ggtgcggaat 141240

ctttctagcg cgcaggacgg cgcgttgccc caactggcca atcggttagg gggagatgcg 141300

gaatcctagg gggggatagg gggtgaggcg gcgaatcggg gccatttggg ggtgctggtc 141360

ggacaacccc tattcgaaag gatccggggt ggcgagtgtt gcggttccgt cgaatgtcct 141420

catagcatcg gcgcgtgatc gcgccgaatt attcttcgca aaaaagagcg tcggcgggtc 141480

gtgtgtccgc gggctttggg gtggaacccg ggtcgctgcg gtggatggtg atcggcgcga 141540

cgggcatgct cggcggcgaa gtggccgccc agctcacggc ccggggcgcc gacccggtgg 141600

gggtcggcag tgcggatctg gacctcaccg acccgcaggc ggtcgccgcg gccgtggccg 141660

acggcggccc cgatgtcgtc gtcaactgcg ccgcctggac cgccgtggac ctggccgaga 141720

ccgaggagga ggcggccctc gccgtcaacg ggacgggagc gggccacctc gcccgggcct 141780

gcgccgccac cggcagccgg ctcctccacg tctccaccga ctacgtcttc cgaggtgccc 141840

cggccgatgc cggacacccc tatgcggagg acgccgaacc cgaccccgcc accgcgtacg 141900

gacgcaccaa gctcgtcggc gagcgcgccg tcctcgccga actccccgcc accgctgccg 141960

tggtgcgcac gtcctggctg tacggacgcg acaacggcgg cttcgtgcac accatggccc 142020

ggctcgcgcg cgagccggga cgcaccgtgg acgtggtcga cgaccagcac ggacagccga 142080

gctggacccc cgatgtcgcg gcccggatca tcgagctcgc cgccctgccc gccgaccggg 142140

cgcacggcgt cttccatgcc accggcgggg gccgcaccac ctggtacgac ctggcccgcg 142200

aggtgttccg gctgaccggc caggacccgg accgggtccg gcgcatcgac agctccgggc 142260

tgcgacgggc ggcggtccgc ccggcatgga gcgttctggg ccatgaccgc tgggccgcca 142320

cggggctcgc cccgatgcgt cactggcgca cggcctcgc ggacgccctc atgggcgacc 142380

ccgtgggcga ccgacttccc gagagtgtga actcccccgg cccgaaaggc tgttgaaggg 142440

tgaaatccct gtcgatagag ggcgcctggc tctatgagcc gctgctccac gacgatgagc 142500

gcggcacgtt cctggaggtg ttccagagcc aggccttcga gctggccacc ggccgccgcc 142560

tcgaactggc ccaggtcaac tgctccgtgt cccgccgcgg cgtcgtgcgc ggcgtccact 142620

tcgccgactt accgcccggc caggccaagt acgtcacctg cgtacgcggc gcggtgcgcg 142680

atgtgatcgt ggacctgcgc accggctcgc ccacctaccg cgcctgggag gccgtcgaac 142740

tcgacgaccg cgaccggcgg gcggtcttcc tctccgaggg cctcggccac gccttccagg 142800
cgatcaccga cgacgccacc gtcgtctacc tgaccacctc gggctacgcc cccggccgtg 142860
agcacggcgt ccacccgctc gacccggagc tgggcatcac ctggcttccc ggcatggaac 142920
cgctgctgtc cccgaaggac gctgtgcccc ccaccctcgc ggtggccgag gcccagggtc 142980
tgctgcccgc gtacgaggac tgcgtacggt acgtgtcctc gctcgccaca ccactcagcg 143040
aggagacccc gtgaaggcac tcgtcctggc gggggatcc ggcacccgcc tgcgcccctt 143100
gacccacacc tcggcgaagc aactcgtgcc cggtggccaa caaacccatc ctcttctacg 143160
tcctggaagg gatcgccgac gcgggcgtca ccgatgtcgg catcatcgtc ggcgacacgg 143220
ccgacgagat cagggcggcc gtcggcgacg gctcccgctt cggcatcagc gtcacctaca 143280
tcccgcagca ccagccgctc ggcctggccc acgccgtgcg catcgcacgg gactggctcg 143340
gcgaggacga cttcgtgatg tacctgggcg acaacttcct gctcggcggg atcagcgagc 143400
agctggagga gttccgcacc cggcggcccg ccgcgcagat catgctcacc cgggtccccg 143460
atccctccgc cttcggcgtc gtcaccctcg acgaggcggg ccgggtcacc ggcctggagg 143520
agaagccgaa gttccccaag agcgatctcg cgctggtcgg cgtgtacttc ttcaccgccg 143580
ccgtgcacga cgccgtggac gccatccagc cctccgcccg cggcgagctg gagatcaccg 143640
aggccctcca gtggctcctc gacaagggcc tcggcatcgc gtcctccacg gtcaacggct 143700
actggaagga caccggcaac gccaccgaca tgctggaggt caaccgcacg gtgctcgaca 143760
ggctgacccc gtactgcgac ggctccgtcg acggcgagag cgaactggtc ggccgggtcg 143820
tcgtcgagga cggcgcggtg atcacccgct cccggatcgt gggccccgcc atcatcggcc 143880
gcggcacccg cgtcgagggc tcctacatcg gcccgttcac ctccgtcggg gcggactgcg 143940
tggtcgtcga cagcgagatc gagtactcca tcgtgctggc cggcgcggcc atcgacggcg 144000
tcggccggat cgaggcgtcc atgatcggcc gtcaggcgca ggtcaccccc gcgccccgca 144060
cgccccaggc ccaccgtctg atcctcggcg accacagcaa ggtgcagatc cgttcatgaa 144120
catcctgatc acgggagcgg ccggcttcat cggctcccac ctcgtacgca cgatcctggg 144180
cccggacaaa ccgctcggcg acgacgtccg cgtcaccgtc ctggacgcgc tgacctacgc 144240
gggcaaccgc gcctccctcg ccgccgtcga ggacgaaccg ggcttcacct tcgtgcacgg 144300
cgacatcacc gacgcgctgc tggtggaccg cctggtggcg gcccacgacg ccgtggtgca 144360
cctggccgcc gagtcgcacg tcgaccgttc gatctggcgg gccgacgcgt tcgtacgcac 144420
caatgtgctc ggcacccaca ccctgctgga ggccgcgctg cggcacggca ccggcccgtt 144480
cgtgcacgtg tcgaccgacg aggtgtacgg ctccggtccg gtcggctcgt ccgtcgagag 144540
cgacccgctg acgcccagct cgccctactc cgcgtccaag gcgtccagtg atctgctggc 144600
cctggcctac caccacaccc acggactcga cgtccccggtg acgcgctgct ccaacaacta 144660